Chemometrics and Intelligent Laboratory Systems 113 (2012) 19–25

Contents lists available at SciVerse ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemolab

# Mixed-level supersaturated design application to a robustness study on an organic synthesis

Alice Croguennoc [a], Marc Frédéric [b], Magalie Claeys-Bruno [c,*], Roger Phan-Tan-Luu [d], Michelle Sergent [c]

[a] Université de Bretagne Occidentale, 3 rue des Archives, CS 93837, F29238 Brest cedex 3, France
[b] Sanofi-Aventis R&D, Site de Vitry-Alfortville, 13 quai Jules Guesde, 94400 Vitry-sur-Seine, France
[c] Aix-Marseille Université, Institut des Sciences Moléculaires de Marseille, ISM2-UMR-CNRS-6263, Equipe AD2EM, Campus St Jérôme, Service D 52, 13397 Marseille Cedex 20, France
[d] LPRAI, 40, boulevard Icard, 13010, Marseille, France

ARTICLE INFO

*Article history:*
Received 24 January 2011
Received in revised form 12 December 2011
Accepted 7 January 2012
Available online 21 January 2012

*Keywords:*
Mixed-level supersaturated designs
Screening
Process robustness
D-optimal design

ABSTRACT

Mixed-level supersaturated designs are designs in which the number of coefficients to estimate is greater than the number of experiments. This type of design is useful in the rapid preliminary investigation of a process with a large number of potentially relevant factors but with only a few of them having important effects. The purpose of this project was to determine the active parameters on a chemical process. A mixed-level supersaturated design of 12 experiments was carried out, allowing us to screen 17 two-level factors and 5 three-level factors. $\chi^2$-optimality of the design was confirmed. In order to verify the results, a 40-run matrix was built. One parameter was identified as very active in both analyses, and several others were suspected to have an effect on the process. Some differences were observed in the results, regarding the detection of the least influent factors. These are likely due to the loss of information generated by the reduction of runs in the supersaturated design, as this kind of design allows an important reduction of the number of experiments. Further studies should be considered to confirm the activity of some factors.



## 1. Introduction

Screening strategies are effective methods to identify active factors, i.e. influent parameters, in a process, and allow the evaluation of the robustness of a process [1]. They are widely used for the analysis of industrial processes. If the number of factors becomes very large, the number of experiments required by classical screening designs may be impractical, especially if the experimental runs are expensive or time consuming (as with crash tests for example). In such cases, if the probability for a factor to be influent is very low (less than 10%, which is called the "sparsity effect"), a supersaturated design may be considered. In this design the number of effects to be estimated is higher than the number of experiments. This kind of design is less expensive and time consuming than classical screening matrices.

First developed in the 1950s by Satterthwaite [2] as a random balance and Booth and Cox [3] in a systematic manner, these designs have recently become increasingly popular. Nevertheless, most studies have focused on two-level supersaturated designs (see for example [4–19]), with several extensions to three-level [20,21] or multi-level [22,23] supersaturated designs. Much more recently mixed-level supersaturated designs were tested [24–28], and were found adapted in cases when the response is based on a polynomial response surface model or in situations where factors are categorical variables.

In this project, realized in the Chemical Development Automation Laboratory of Sanofi-Aventis in Vitry sur Seine, a mixed-level supersaturated (asymmetric) design was carried out in order to estimate the robustness of an organic synthesis. A 12-run supersaturated design was constructed, allowing us to screen 17 two-level factors and 5 three-level factors. As this mixed-level supersaturated design is not conventional and has seldom been studied, a 40-run screening matrix (D-optimal, [29]) was built in order to compare the results. This matrix was a complement of the supersaturated design: all of the experiments of the supersaturated design were included in this D-optimal matrix.

## 2. Application of a supersaturated design to an organic synthesis

### 2.1. Synthesis

The study applies to the industrial synthesis of PBA salt, an intermediate of synthesis of an anti-cancer drug. A mixed-level supersaturated design was carried out in order to evaluate the influent effects on the process, and more especially the purity of the salt obtained.

#### 2.1.1. Reaction

The PBA salt is made using the reaction presented in Fig. 1.

* Corresponding author. Tel.: +33 491288186.
*E-mail address:* m.claeys-bruno@univ-cezanne.fr (M. Claeys-Bruno).

0169-7439/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.chemolab.2012.01.005

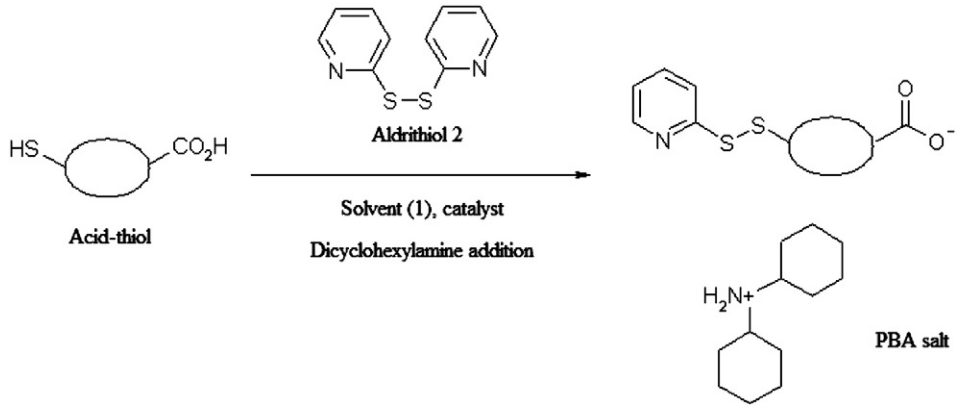


**Fig. 1.** Synthesis reaction.

*2.1.2. Protocol*

2,2′-Dithiopyridine (1.5 eq) was dissolved in a solvent (1) (5 V/dithiopyridine) and stirred for 10 min. The solution was cooled down to 0 ± 3 °C, and a catalyst added (0.02 eq). A solution of acid-thiol (1 eq) was then added at a temperature below 3 °C. The reaction mixture was stirred for 5 h at 0 ± 3 °C and then concentrated under reduced pressure to half its initial volume $V_0$. The concentrated mixture was stirred for 1 h at 0 ± 3 °C and the precipitated by-product (thiopyridone) was isolated by filtration. Dicyclohexylamine (1.2 eq) was then added to the mother liquor at 15 ± 3 °C, and the mixture cooled down to 0 ± 5 °C. The mixture was then stirred for 5 h at 0 ± 3 °C. The solid product (dicyclohexylamine salt) was isolated by filtration, washed with solvent (1) (2 × 12.5 V) and dried under reduced pressure at 40 °C. The purity of the dried product was determined by HPLC.

This basic protocol was adapted to each experiment, in order to allow the variations of the studied factors.

*2.1.3. Equipment*

All experiments of the supersaturated design and some of the D-optimal design were carried out using a multi-reactors Polyblock H.E.L®, piloted by WinIso software, and reactors volumes were either of 150 mL or 100 mL capacities. The Polyblock can receive 4 reactors independently controlled in terms of temperature and stirring. Some of the experiments of the 40-run were done on an Auto-MATE H.E.L®, also piloted by WinIso, using 100 mL and 50 mL reactors. The purity of the obtained salt was then determined by an HPLC analysis, using a Waters® Alliance equipped by an XTerra RP8 150 * 4.6 mm–3.5 μm column. Data acquisition was made via Waters Empower® software. Results were obtained thanks to an external calibration, by comparison of the sample's chromatogram to a known standard, considered 100% pure. 20 μL of sample solution were eluted by a mixture of acetonitrile and water at 1 mL/min flow for 35 min, and analysed in UV detection, at 290 nm.

*2.2. Experimental design*

*2.2.1. Factors and domain of interest*

In this study, 22 factors were studied, and one response (the purity of the isolated salt) was analysed. The factors were chosen based on the reference protocol. Three different suppliers were used for each of the reagents and three different catalysts were tested. Factors and levels are shown in Table 1 in which eq is the molar equivalent in acid-thiol, V is volume in mL per g of dithiopyridine and $V_0$ is the initial volume.

One response, the purity of the PBA salt, was analysed using the supersaturated design.

*2.2.2. Construction of the supersaturated design*

The mixed-level supersaturated design ($N = 12$) was constructed according to Yamada [24] using $C_2$ and $D_3$ designs (constructed by lexicographical enumeration and computer search respectively) in association with the generating designs $T^2{}_2$ and $T^2{}_3$, also proposed by Yamada [24]. These matrices are presented below.

$$C_2 = \begin{pmatrix} 1&1&1&1&1&1&1&1&1&1\\ 1&1&1&1&2&2&2&2&2&2\\ 1&2&2&2&1&1&1&2&2&2\\ 2&1&2&2&1&2&2&1&1&2\\ 2&2&1&2&2&1&2&1&2&1\\ 2&2&2&1&2&2&1&2&1&1 \end{pmatrix} \qquad D_3 = \begin{pmatrix} 1&1&1&1&1\\ 2&1&2&3&3\\ 3&2&3&3&1\\ 1&2&2&2&2\\ 2&3&3&1&2\\ 3&3&1&2&3 \end{pmatrix}$$

$$T^2{}_2 = \begin{pmatrix} 0&0\\ 0&1 \end{pmatrix} \qquad T^2{}_3 = \begin{pmatrix} 0&0\\ 1&2 \end{pmatrix}$$

A matrix $C = T^2{}_2 \oplus C_2$ is then generated, in which the operator $\oplus$ determines the $((i-1)lm+u,\ (j-1)p+v)$ element of the matrix C by $mod(t_{ij}+c_{uv}-1,l)+1$ , with $l$ number of levels of $T^2{}_2$ (in this case $l = 2$), $m$ number of levels of $T^2{}_3$ (here $m = 3$), $p$ number of

**Table 1**
Factors and experimental domain of interest.

| | Factors | Number of levels | Level 1 | Level 2 | Level 3 |
|---|---|---|---|---|---|
| U1 | Dithiopyridine quantity | 2 | 1.47 eq | 1.53 eq | |
| U2 | Solvent quantity | 2 | 4.5 V | 5.5 V | |
| U3 | Stirring time 1 | 2 | 5 min | 15 min | |
| U4 | Temperature 1 | 2 | −3 °C | +3 °C | |
| U5 | Acid-thiol quantity | 2 | 0.98 eq | 1.02 eq | |
| U6 | Acid-thiol adding time | 2 | 5.5 min | 16.5 min | |
| U7 | Catalyst quantity | 2 | 0.01 eq | 0.03 eq | |
| U8 | Stirring time 2 | 2 | 4.5 h | 5.5 h | |
| U9 | Concentration: end volume | 2 | 0.4 $V_0$ | 0.6 $V_0$ | |
| U10 | Rinsing 1 | 2 | 0.14 V | 0.16 V | |
| U11 | Washing 1 | 2 | 0.48 V | 0.53 V | |
| U12 | Rinsing 2 | 2 | 0.14 V | 0.16 V | |
| U13 | Temperature 2 | 2 | 12 °C | 18 °C | |
| U14 | Dicyclohexylamine quantity | 2 | 1.17 eq | 1.22 eq | |
| U15 | Dicyclohexylamine dosing time | 2 | 7.5 min | 22.5 min | |
| U16 | Temperature 3 | 2 | −3 °C | +3 °C | |
| U17 | Washing 2 | 2 | 1.43 V | 1.58 V | |
| U18 | Dithiopyridine supplier | 3 | Alfa Aesar® 98% | Aldrich® 98% | Acros® 98% |
| U19 | Solvent supplier | 3 | SDS® 99.8% | Acros® 99.5% | Prolabo® 99.9% |
| U20 | Catalyst type | 3 | Catalyst 1 | Catalyst 2 | Catalyst 3 |
| U21 | Dicyclohexylamine supplier | 3 | Acros® 99% | Alfa Aesar® 98% | Aldrich® 99% |
| U22 | Stirring time 3 | 3 | 15 h | 18 h | 23 h |

**Table 2**
Supersaturated design.

| Exp | X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 | X9 | X10 | X11 | X12 | X13 | X14 | X15 | X16 | X17 | X18 | X19 | X20 | X21 | X22 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 2 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 2 | 3 | 3 |
| 3 | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 2 | 2 | 2 | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 3 | 2 | 3 | 3 | 1 |
| 4 | 2 | 1 | 2 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 2 | 1 | 2 | 2 | 1 | 2 | 2 | 2 | 2 |
| 5 | 2 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 2 | 1 | 2 | 2 | 3 | 3 | 1 | 2 |
| 6 | 2 | 2 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 2 | 1 | 2 | 2 | 1 | 3 | 3 | 1 | 2 | 3 |
| 7 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 8 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 3 | 2 | 3 | 1 | 1 |
| 9 | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 2 | 2 | 2 | 1 | 3 | 1 | 1 | 2 |
| 10 | 2 | 1 | 2 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 1 | 2 | 3 | 3 | 3 | 3 |
| 11 | 2 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 3 | 1 | 1 | 2 | 3 |
| 12 | 2 | 2 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 1 | 2 | 3 | 1 |

columns of $C_2$, and $t_{ij}$ and $c_{uv}$ positions $(i,j)$ and $(u,v)$ of the elements of $T^2{}_2$ and $C_2$ respectively. In the same way, a matrix $D = T^2{}_3 \oplus D_3$ is generated, in which $\oplus$ determines the $((i-1)lm+u,(j-1)q+v)$ element of the matrix D by $mod(tij+duv-1,m)+1$, with $q$ number of columns of $D_3$, and $t_{ij}$ and $d_{uv}$ positions $(i,j)$ and $(u,v)$ of the elements of $T^2{}_3$ and $D_3$ respectively. A $2^{20}3^{10}//12$ candidate matrix was then obtained by the juxtaposition of the matrix C and the matrix D. By deleting unnecessary columns, we obtain a $2^{17}3^5//12$ supersaturated design. This resulting mixed-level supersaturated design is presented in Table 2.

### 2.2.3. Construction of the D-optimal design

The complementary 40-run matrix is a traditional D-optimal matrix which includes the supersaturated design experiments [29]. A conventional screening design $2^{17}3^5$ would contain 36 runs, to which we should add the 12 experiments already carried out in the supersaturated design: 48 distinct experiments are obtained. In order to reduce the number of experiments, the Fedorov algorithm, used by NemrodW® [30], selected a design with fewer experiments and a satisfactory quality, including all the experiments from the supersaturated design. In this case study, the inflation factor was chosen as the acceptability criterion [29]. This factor must be equal or inferior to 2 (in this case) to be accepted. A 40-run matrix is obtained via this algorithm, as shown in Fig. 2. This matrix is presented in Table 3, in which the experiences 1 to 12 are the ones from the supersaturated design. The experiences marked with an apostrophe are repetitions. The lines in italics with a 'P' in the second column are the experiments done using the Polyblock, and the others are the ones carried out using the Auto-MATE, at a different scale.

### 2.2.4. Quality criterion $\chi^2$

Several quality criteria have been proposed in the bibliography, such as the degree of saturation [24], the $\chi^2$ criterion [24], the $E(f_{NOD})$ criterion [27,25,31,32]. In most of these criteria, the point is to evaluate the non-orthogonality of the matrix. In order to simplify the analysis, only one criterion was studied: the $\chi^2$-optimality criterion, proposed by Yamada [26] in the publication used for the construction of the mixed-level supersaturated design. This criterion should be minimized to obtain the best matrix possible in term of orthogonality. It is presented in Eq. (1):

$$\chi^2(c,d) = \sum_{a\in\{1,\ldots,l\}} \sum_{b\in\{1,\ldots,m\}} \frac{\left(n^{ab}(c,d) - n/(lm)\right)^2}{n/(lm)} \tag{1}$$

In which $c$ is a column from the matrix C presented above, $d$ a column from the matrix D, $l$ the number of levels of the matrix C and $m$ the number of levels of the matrix D. $n$ is the number of runs, and $n^{ab}(c,d)$ the number of lines of the matrix $n\times 2$ formed by $(c,d)$ which present as modalities $(a,b)$.

Several variations of the $\chi^2$ optimality criterion have been calculated, for the full constructed supersaturated design and each of its matrices C and D.

Let C be a $l$ levels matrix, including $n$ runs and $p$ columns, $(c_i,c_j)$ columns of C, and D a $m$ levels matrix, including $n$ runs and $q$ columns, $(d_i, d_j)$ columns of D. An average $\chi^2$ is calculated for each of these matrices and for all couples $(c,d)$.

$$\text{Average } \chi^2 \text{ of the matrix C} \qquad \text{ave}\chi^2_{l,l} = \sum_{1\le i\le j\le p} \chi^2\left(c_i,c_j\right) \Big/ \binom{p}{2} \tag{2}$$

$$\text{Average } \chi^2 \text{ of the matrix } D \qquad \text{ave}\chi^2_{m,m} = \sum_{1\le i\le j\le q} \chi^2\left(d_i,d_j\right) \Big/ \binom{q}{2} \tag{3}$$

$$\text{Average } \chi^2 \text{ of all couples}(c,d) \qquad \text{ave}\chi^2_{l,m} = \sum_{1\le i\le p} \sum_{1\le j\le q} \chi^2\left(c_i,d_j\right) \Big/ (pq) \tag{4}$$

Let $\chi^2$ sum be the sum of all the $\chi^2$ calculated in the previous equations (Eqs. (2) to (4)), and Ds the degree of saturation.

$$\text{Saturation degree} \qquad Ds = \frac{(l-1)p + (m-1)q}{n-1} \tag{5}$$

The next criterion, $\chi^2$-eff, measures the capacity of the matrix to reach $\chi^2$ optimality:

$$\chi^2\text{-efficiency} \qquad \chi^2_{eff} = \frac{Ds(Ds-1)n(n-1)/2}{\chi^2_{sum}} \tag{6}$$

All of these criteria have been calculated, and are presented in Table 4. The design is optimal in terms of orthogonality if the criteria Max and Ave are minimised, and if the criterion $\chi^2$-eff is as close as 1 as possible. In the presented case, the quality of the constructed supersaturated design was verified, as we have $\chi^2$-eff = 0.74, which is satisfactory.

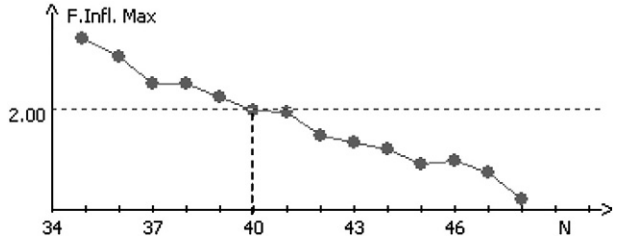


**Fig. 2.** Inflation factor.

**Table 3**
D-optimal matrix. The lines in italics with a 'P' in the second column correspond to the experiments done using the Polyblock, and the others are the ones carried out using the AutoMATE, at a different scale. The repeated experiments are marked with an apostrophe.

| Exp | | X1 | X2 | X3 | X4 | X5 | X6 | X7 | X8 | X9 | X10 | X11 | X12 | X13 | X14 | X15 | X16 | X17 | X18 | X19 | X20 | X21 | X22 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *1* | *P* | *1* | *1* | *1* | *1* | *1* | *1* | *1* | *1* | *1* | *1* | *1* | *1* | *1* | *1* | *1* | *1* | *1* | *1* | *1* | *1* | *1* | *1* |
| *2* | *P* | *1* | *1* | *1* | *1* | *2* | *2* | *2* | *2* | *2* | *2* | *1* | *1* | *1* | *1* | *2* | *2* | *2* | *2* | *1* | *2* | *3* | *3* |
| *3* | *P* | *1* | *2* | *2* | *2* | *1* | *1* | *1* | *2* | *2* | *2* | *1* | *2* | *2* | *2* | *1* | *1* | *1* | *3* | *2* | *3* | *3* | *1* |
| *4* | *P* | *2* | *1* | *2* | *2* | *1* | *2* | *2* | *1* | *1* | *2* | *2* | *1* | *2* | *2* | *1* | *2* | *2* | *1* | *2* | *2* | *2* | *2* |
| *5* | *P* | *2* | *2* | *1* | *2* | *2* | *1* | *2* | *1* | *2* | *1* | *2* | *2* | *1* | *2* | *2* | *1* | *2* | *2* | *3* | *3* | *1* | *2* |
| *6* | *P* | *2* | *2* | *2* | *1* | *2* | *2* | *1* | *2* | *1* | *1* | *2* | *2* | *2* | *1* | *2* | *2* | *1* | *3* | *3* | *1* | *2* | *3* |
| *7* | *P* | *1* | *1* | *1* | *1* | *1* | *1* | *1* | *1* | *1* | *1* | *2* | *2* | *2* | *2* | *2* | *2* | *2* | *2* | *2* | *2* | *2* | *2* |
| *8* | *P* | *1* | *1* | *1* | *1* | *2* | *2* | *2* | *2* | *2* | *2* | *2* | *2* | *2* | *2* | *1* | *1* | *1* | *3* | *2* | *3* | *1* | *1* |
| *9* | *P* | *1* | *2* | *2* | *2* | *1* | *1* | *1* | *2* | *2* | *2* | *2* | *1* | *1* | *1* | *2* | *2* | *2* | *1* | *3* | *1* | *1* | *2* |
| *10* | *P* | *2* | *1* | *2* | *2* | *1* | *2* | *2* | *1* | *1* | *2* | *1* | *2* | *1* | *1* | *2* | *1* | *1* | *2* | *3* | *3* | *3* | *3* |
| *11* | *P* | *2* | *2* | *1* | *2* | *2* | *1* | *2* | *1* | *2* | *1* | *1* | *1* | *2* | *1* | *1* | *2* | *1* | *3* | *1* | *1* | *2* | *3* |
| 11′ | | 2 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 3 | 1 | 1 | 2 | 3 |
| *12* | *P* | *2* | *2* | *2* | *1* | *2* | *2* | *1* | *2* | *1* | *1* | *1* | *1* | *1* | *2* | *1* | *1* | *2* | *1* | *1* | *2* | *3* | *1* |
| 13 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 3 | 3 | 2 | 2 | 3 |
| *14* | *P* | *2* | *2* | *2* | *1* | *1* | *1* | *2* | *1* | *2* | *1* | *2* | *2* | *2* | *1* | *1* | *1* | *1* | *2* | *2* | *3* | *3* | *2* |
| *15* | *P* | *1* | *2* | *1* | *2* | *2* | *2* | *2* | *1* | *1* | *1* | *2* | *1* | *1* | *2* | *2* | *2* | *2* | *3* | *3* | *3* | *3* | *2* |
| 16 | | 2 | 1 | 2 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 2 | 2 | 3 |
| 17 | | 1 | 1 | 1 | 2 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 2 | 1 | 1 | 1 | 3 | 1 | 1 | 2 | 1 |
| 18 | | 2 | 2 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 2 | 1 | 2 | 1 | 1 | 1 | 1 | 3 | 2 | 1 | 2 |
| *19* | *P* | *1* | *2* | *2* | *1* | *1* | *2* | *2* | *2* | *1* | *2* | *1* | *1* | *2* | *1* | *2* | *1* | *1* | *3* | *1* | *2* | *1* | *2* |
| *20* | *P* | *2* | *1* | *2* | *2* | *1* | *2* | *1* | *2* | *1* | *1* | *2* | *1* | *2* | *1* | *2* | *1* | *1* | *1* | *3* | *1* | *2* | *1* |
| *21* | *P* | *1* | *1* | *2* | *1* | *2* | *2* | *1* | *1* | *2* | *2* | *2* | *2* | *1* | *1* | *2* | *1* | *2* | *2* | *1* | *3* | *1* | *1* |
| 22 | | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 3 | 3 |
| *23* | *P* | *1* | *2* | *2* | *2* | *2* | *1* | *1* | *2* | *2* | *1* | *1* | *1* | *2* | *2* | *1* | *1* | *2* | *2* | *1* | *1* | *3* | *3* |
| *24* | *P* | *2* | *1* | *1* | *1* | *2* | *2* | *2* | *2* | *2* | *1* | *1* | *1* | *2* | *2* | *1* | *2* | *1* | *1* | *2* | *3* | *1* | *1* |
| 25 | | 2 | 2 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 1 | 2 | 1 | 3 | 2 | 3 |
| 26 | | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 2 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 2 | 3 | 2 | 3 |
| *27* | *P* | *2* | *1* | *2* | *2* | *1* | *2* | *1* | *2* | *1* | *1* | *2* | *1* | *1* | *1* | *1* | *2* | *1* | *2* | *1* | *2* | *3* | *2* |
| 28 | | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 1 | 3 | 2 | 1 | 2 | 2 |
| 28′ | | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 1 | 3 | 2 | 1 | 2 | 2 |
| *29* | *P* | *2* | *2* | *1* | *2* | *1* | *2* | *1* | *1* | *2* | *2* | *1* | *1* | *1* | *1* | *1* | *1* | *2* | *2* | *3* | *2* | *1* | *1* |
| *30* | *P* | *1* | *2* | *2* | *2* | *2* | *1* | *1* | *2* | *2* | *1* | *1* | *1* | *1* | *1* | *1* | *2* | *1* | *3* | *2* | *2* | *1* | *1* |
| 31 | | 2 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 3 | 1 | 2 | 2 |
| *32* | *P* | *2* | *2* | *2* | *1* | *1* | *1* | *2* | *1* | *2* | *1* | *2* | *1* | *1* | *2* | *2* | *2* | *2* | *1* | *1* | *2* | *2* | *1* |
| 33 | | 1 | 2 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 1 |
| 34 | | 2 | 1 | 2 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 1 | 2 |
| 35 | | 2 | 2 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 3 | 2 | 1 | 3 | 1 |
| 36 | | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 1 | 2 | 3 | 1 | 3 | 1 |
| 37 | | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 2 | 3 | 2 | 3 | 1 | 3 |
| 38 | | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 2 | 1 | 2 | 2 | 1 | 1 | 1 | 1 | 3 | 2 | 3 | 3 |
| 39 | | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 3 | 1 | 3 | 2 | 2 |
| 39′ | | 2 | 2 | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 1 | 3 | 1 | 3 | 2 | 2 |
| 40 | | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 1 | 2 | 1 | 1 | 1 | 3 | 3 | 2 | 2 |

## 3. Results and discussion

One response Y was studied by the designs: the purity of the obtained PBA salt, measured by HPLC analysis. The results are presented in Table 5. Y is expressed in %. However, some of the results appear to be above 100% in purity, which is impossible. This can be explained by the calculation of the response: measurements were made thanks to an external standard that was considered as a hundred % pure. The obtained results show that several experiments allowed us to synthesize a better quality salt.

### 3.1. Supersaturated design

Data processing was carried out in two steps: first a step-wise regression was performed, and then all subset regressions were used for a number of factors from two to six. Four criteria were studied in order to determine the number of necessary variables in the model: $R^2$ (correlation coefficient), AIC (Akaike Information Criterion [33]), BIC (Bayesian Information Criterion [34]) and $s^2$ (residual variance). $R^2$ should be maximised, and the other criteria minimised. AIC and BIC are useful tools for the model selection: they allow us to evaluate the model adequacy and grant the comparison between several

**Table 4**
Quality criteria from Yamada [24].

| | Sum ($\chi^2$) | Ds | Max | Ave | $\chi^2$-eff |
|---|---|---|---|---|---|
| C (12×17) | 88 | 1.55 | 1.33 | 0.65 | 0.63 |
| (C,D) | 186 | 2.45 | 6 | 2.4 | 0.65 |
| Full | 316 | 2.45 | 6 | | 0.74 |

**Table 5**
Results: purity of the salt in %.

| SSD | Y | D-opt | Y | D-opt | Y |
|---|---|---|---|---|---|
| 1 | 97.59 | 13 | 99.28 | 28 | 101.43 |
| 2 | 86.68 | 14 | 100.13 | 28′ | 101.15 |
| 3 | 98.39 | 15 | 101.13 | 29 | 79.09 |
| 4 | 100.26 | 16 | 90.82 | 30 | 96.83 |
| 5 | 90.19 | 17 | 99.15 | 31 | 100.49 |
| 6 | 99.01 | 18 | 98.71 | 32 | 98.8 |
| 7 | 80.33 | 19 | 99.96 | 33 | 87.22 |
| 8 | 94.3 | 20 | 92.74 | 34 | 100.34 |
| 9 | 94.65 | 21 | 82.98 | 35 | 103.67 |
| 10 | 95.66 | 22 | 102.39 | 36 | 104.09 |
| 11 | 100.8 | 23 | 79.74 | 37 | 98.86 |
| 11′ | 102.33 | 24 | 95.62 | 38 | 99.53 |
| 12 | 97.73 | 25 | 98.27 | 39 | 102.55 |
| | | 26 | 101.45 | 39′ | 98.38 |
| | | 27 | 81.72 | 40 | 101.49 |

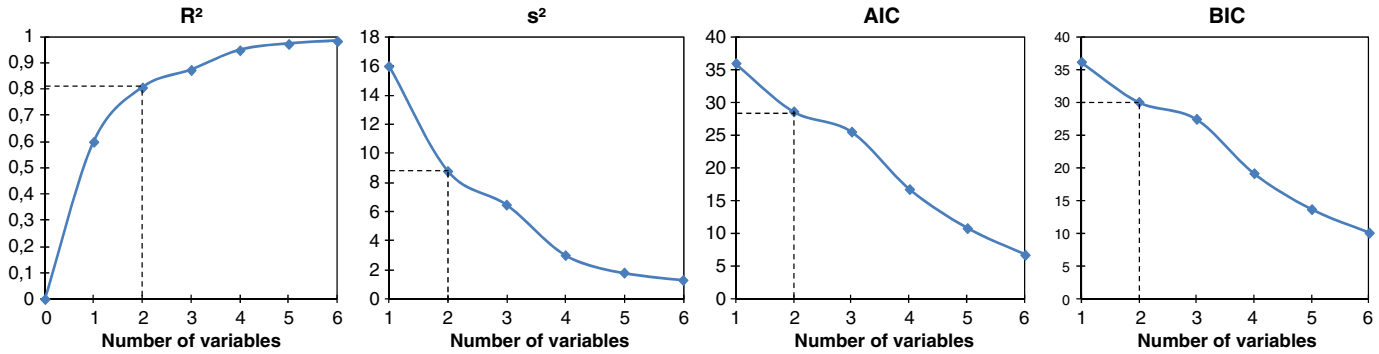


**Fig. 3.** Evolution of the criteria in function of the number of variables.

experimental models. They are helpful in the choice of the right number of parameters for the model. The following equations (Eqs. (8) and (9)) show how both criteria are calculated, in which $k$ is the number of the estimated parameters of the model, $n$ the number of observations and $L$ the maximum probability density for the estimated model:

$$\mathrm{AIC} = -2\ln(L) + 2k \tag{8}$$

$$\mathrm{BIC} = -2\ln(L) + k\ln(n) \tag{9}$$

Fig. 3 gives the progress of these criteria based on the number of variables regressed. In the case of the $R^2$ criterion, the introduction of the first variable leads to an important raise. The introduction of a second variable leads to a weaker increase, and the next ones do not induce any significant raise of the criterion $R^2$. Thus, this criterion indicates that a two-variable model seems acceptable in this analysis. The other criteria shown on Fig. 3 confirm that a two-variable model is acceptable, as a rupture can be seen in the curves after the second variable is added.

The mapping of the variables is then studied [30] (Fig. 4). In this figure, the selected variables appear in black for each proposed model, from the best model (in terms of criteria $R^2$ only) in the first line above, that we can call the first model, to the eight model (last line of the figure).

Two variables appear regularly selected: b18B and b1A, which means that the parameter X18 is predominant, followed by the factor X1. These two parameters are confirmed by the stepwise regression. These figures indicate that two variables seem to be a reasonable choice to explain the variation of the response. Thus, two factors are detected as influent on the process, U18 (dithiopyridine supplier) and U1 (dithiopyridine quantity).

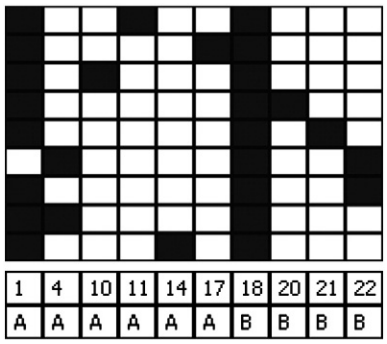


**Fig. 4.** Mapping with 3 variables.

### 3.2. D-optimal matrix

Data analysis of the D-optimal matrix is conventional: first the variance is studied, then the residues and the distribution of the data, and finally the factor effect plot is analysed.

The variance study is shown in the ANOVA table (Table 6). The ANOVA allows us to determine if the response is well explained by the postulated model. In the present case, there is a 0.6% probability that the null hypothesis is refused. We conclude then that the

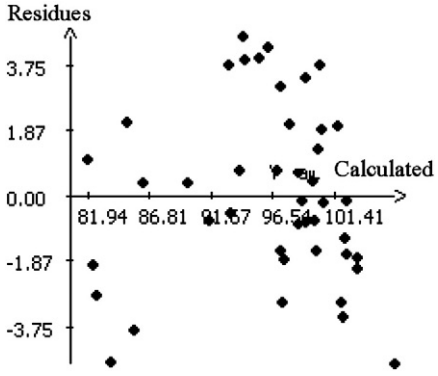


**Fig. 5.** Residues repartition.

**Table 6**
Variance analysis on the response Y.

| Variation source | Square sum | Freedom degrees | Mean square | Report | Signif |
|---|---|---|---|---|---|
| Regression | $1.73E^{+3}$ | 27 | 64.19 | 3.62 | 0.586 ** |
| Residues | $2.66E^{+2}$ | 15 | 17.74 | | |
| Validity | $2.56E^{+2}$ | 12 | 21.35 | 6.47 | 7.5 |
| Error | 9.90 | 3 | 3.30 | | |
| Total | $2.00E^{+3}$ | 42 | | | |

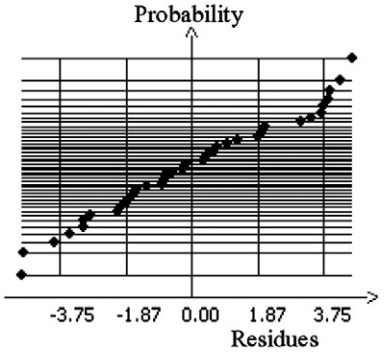


**Fig. 6.** Daniel's plot.

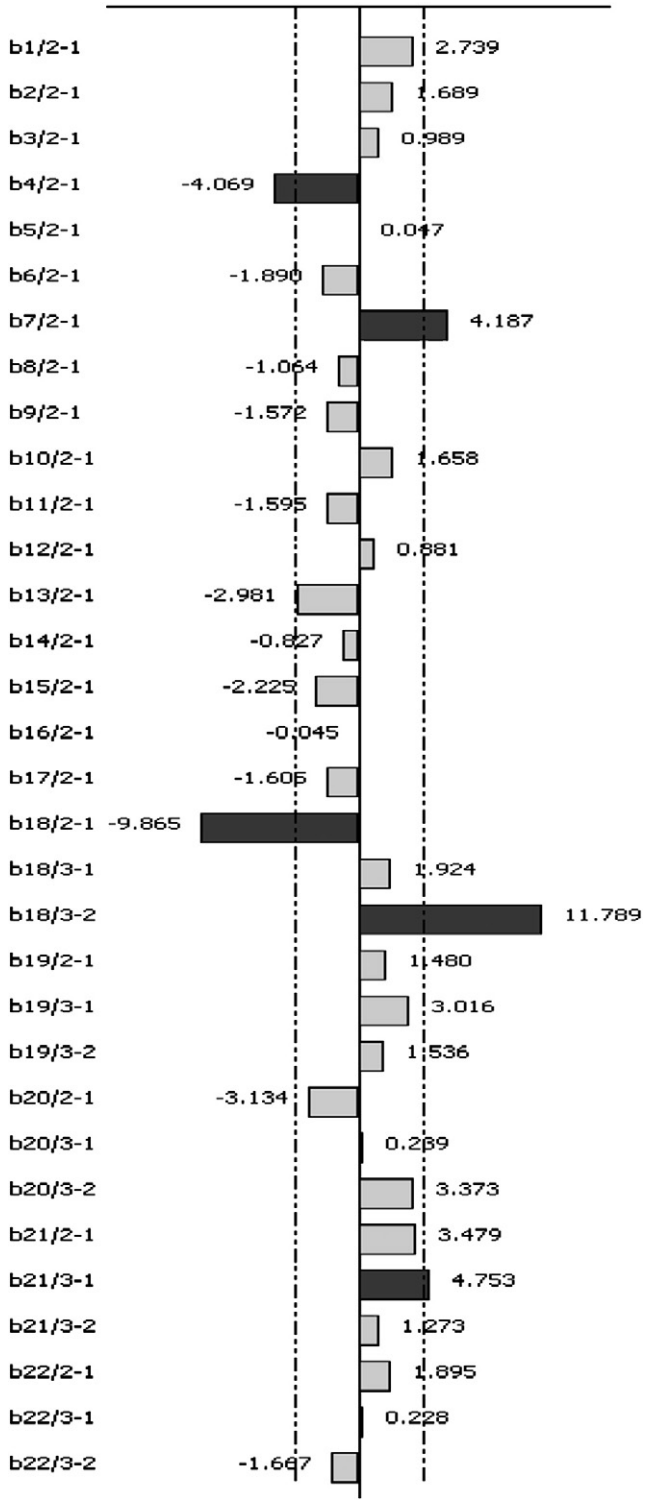


**Fig. 7.** Factor effect plot of the D-optimal design.

response is well-explained by the postulated model. In the same manner, the analysis based on a Fisher test indicates that there is no lack of fit detected. Then, there is nothing in the variance analysis that permits us to reject the postulated model.

Next, the residues are low and homoscedastic (Fig. 5), and they follow a normal plot [35] (Fig. 6).

Finally, the analysis of the factor effect plot (Fig. 7) indicates that the most influent factor is U18 (dithiopyridine supplier), and two other parameters have little influence (U4, temperature 1 and U7, catalyst quantity). The significance of another factor, U21 (dicyclohexylamine supplier), is debatable.

To sum up, four factors are detected as actually active: U18 in a large way, and U4, U7 and U21 to a lesser extent.

## 4. Conclusion

The results of this study are quite encouraging. Firstly, the supersaturated design permitted to confirm the robustness of the process protocol studied in only 12 experiments.

Secondly, the same most active parameter (U18, dithiopyridine supplier) was found with both matrices, mixed-level supersaturated and D-Optimal. This reagent origin effect cannot be explained, as the analyses certificates provided by all three suppliers indicate that the products are more than 99.5% pure. In the analysis of the 40-run matrix, a few less influent factors were found, but this difference can be explained by the loss of information due to the supersaturated design, entailing the reduction of experiments. Moreover, in the present study, the D-optimal analysis could indicate that the sparsity effect is not ideal, as four active factors are detected, which is more than the 10% limit required. As supersaturated designs are used as preliminary studies, and permit to identify only the most important factors, further analyses of the process with conventional matrices should be considered. Regardless, the use of this kind of matrices permit a significant gain of time, as in this study, a 12-run supersaturated design allowed us to carry out the same robustness study as a 40-run conventional design.

## Acknowledgements

We would like to thank Denis Prat and Diké Alozié (Sanofi-Aventis) for the fruitful discussion.

## References

[1] F. Rais, A. Kamoun, M. Chaabouni, M. Claeys-Bruno, R. Phan-Tan-Luu, M. Sergent, Supersaturated design for screening factors influencing the preparation of sulphated amides of olive pomace oil fatty acids, Chemometrics and Intelligent Laboratory Systems 99 (2009) 71–78.
[2] F. Satterwaithe, Random balance experimentation (with discussion), Technometrics 1 (1959) 111–137.
[3] K.H.V. Booth, D.R. Cox, Some systematic supersaturated designs, Technometrics 4 (1962) 489–495.
[4] X. Lu, Y. Meng, A new method in the construction of two-level supersaturated design, Journal of Statistical Planning and Inference 86 (2000) 229–238.
[5] T. Beth, D. Jungnickel, H. Lenz, Design Theory, Cambridge, 1999.
[6] N.A. Butler, R. Mead, K.M. Eskridge, S.G. Gilmour, A general method of constructing $E(s^2)$-optimal supersaturated designs, Journals of the Royal Statistical Society B 63 (2001) 621–632.
[7] C.S. Cheng, E(s2)-optimal supersaturated designs, Statistica Sinica 7 (1997) 929–939.
[8] L.-Y. Deng, D.K.J. Lin, J. Wang, A resolution rank criterion for supersaturated designs, Statistica Sinica 9 (1999) 605–610.
[9] W.W. Li, C.F.J. Wu, Columnwise-pairwise algorithms with applications to the construction of supersaturated designs, Technometrics 39 (1997) 171–179.
[10] D.K.J. Lin, A new class of supersaturated designs, Technometrics 35 (1993) 28–31.
[11] D.K.J. Lin, Generating systematic supersaturated designs, Technometrics 37 (1995) 213–225.
[12] M.-Q. Liu, F.J. Hickernell, $E(s^2)$-optimality and minimum discrepancy in 2-level supersaturated designs, Statistica Sinica 12 (2002) 931–939.
[13] B. Tang, C.F.J. Wu, A method for constructing supersatured designs and its E(s2) optimality, Canadian Journal of Statistics 25 (1997) 191–201.
[14] N.-K. Nguyen, An algorithmic approach to constructing supersaturated designs, Technometrics 38 (1996) 69–73.
[15] M.-Q. Liu, R. Zhang, Construction of E(s2)optimal supersaturated design using cyclic BIBDs, Journal of Statistical Planning and Inference 91 (2000) 139–150.
[16] C.F.J. Wu, Construction of supersaturated designs through partially aliased interactions, Biometrika 80 (1993) 661–669.
[17] S. Yamada, D.K.J. Lin, Supersaturated designs including an orthogonal base, Canadian Journal of Statistics 25 (1997) 203–213.
[18] M. Claeys-Bruno, M. Dobrijevic, R. Phan-Tan-Luu, M. Sergent, A new class of supersaturated design: application to a sensitivity study of a photochemical model, Chemometrics and Intelligent Laboratory System 95 (2009) 86–93.

[19] M. Claeys-Bruno, M. Dobrijevic, R. Cela, R. Phan-Tan-Luu, M. Sergent, Supersaturated designs for computer experiments: a useful tool, Application at the determination of key reactions in photochemical models applied to the production of hydrocarbons on the atmosphere of Titan. Chemometrics and Intelligent Laboratory Systems 2 (2011) 137–146.
[20] S. Yamada, Y.T. Ikebe, H. Hashiguchi, N. Niki, Construction of three-level supersaturated design, Journal of Statistical Planning and Inference 81 (1999) 183–193.
[21] S. Yamada, D.K.J. Lin, Three-level supersaturated design, Statistics & Probability letters 45 (1999) 31–39.
[22] K.-T. Fang, D.K.J. Lin, C.-X. Ma, On the construction of multi-level supersaturated designs, Journal of statistical planning and inference 86 (2000) 239–252.
[23] K.T. Fang, G.N. Ge, M.Q. Liu, H. Qin, Construction of minimum generalized aberration designs, Metrika 57 (2003) 37–50.
[24] S. Yamada, M. Matsui, T. Matsui, D.K.J. Lin, T. Takahashi, A general construction method for mixed level supersaturated design, Computational Statics & Data Analysis 50 (2006) 254–265.
[25] P.F. Li, M.-Q. Liu, R.C. Zhang, Some theory and the construction of mixed level supersaturated designs, Statistics & Probability letters 69 (2004) 105–116.
[26] S. Yamada, T. Matsui, Optimality of mixed-level supersaturated designs, Journal of Statistical Planning and Inference 104 (2002) 459–468.
[27] K.-T. Fang, D.K.J. Lin, M.-Q. Liu, Optimal mixed-level supersaturated design and computer experiment, Metrika 58 (2003) 279–291.
[28] K.-T. Fang, G. Ge, M.-Q. Liu, H. Qin, Combinatorial constructions for optimal supersaturated designs, Discrete Mathematics 279 (2004) 191–202.
[29] R. Phan-Tan-Luu, M. Sergent, Non classical experimental design, Comprehensive Chemometrics, Elsevier, 2009 Chapter 1.14,453-499.
[30] D. Mathieu, J. Nony, R. Phan-Tan-Luu, Logiciel NEMRODW, LPRAI, Marseille, 2009.
[31] H. Xu, Minimum Moment Aberration for non-regular designs and supersaturated designs, Statistica Sinica 13 (2003) 691–708.
[32] M. Ai, K. Fang, S. He, E($\chi^2$)-optimal mixed-level supersaturated designs, Journal of Statistical Planning and Inference 137 (2007) 306–316.
[33] H. Akaike, A new look at the statistical identification model, IEEE Transactions on Automatic Control 19 (1974) 716–723.
[34] G. Schwarz, Estimating the dimension of a model, The Annals of Statistics 6 (1978) 461–464.
[35] D. Cuthbert, Application of statistics in industrial experimentation, J. Wiley and Sons, New York, London, Sydney, Toronto, 1976.