Proceedings of the World Congress on Engineering 2007 Vol II
WCE 2007, July 2 - 4, 2007, London, U.K.

Objective Methods for Analysing Unreplicated Factorial Designs

Nuno Costa and Zulema Lopes Pereira, *Member, IAENG*

Abstract — There is not a method for analysing unreplicated factorial designs that performs well for various configurations of number and size of active effects. Moreover, the most popular tool that has been applied is informal and subjective. To overcome these drawbacks, this paper suggests a multiple testing to help practitioners making objective decisions and identifying active effects with a minimum number of additional runs for a wide range of active effects configurations. The selected methods are efficient and easy to implement by practitioners, including those who do not have a profound knowledge of statistics. Two examples from the literature exemplify and justify the proposed approach.

Index Terms — Design of experiments, Error rate, Location Effects, Screening.

I. Introduction

Managers and engineers of today's modern industrial world have placed an increased emphasis on achieving breakthrough improvements in productivity and quality of processes and products through the application of Design of Experiments (DoE) and other statistical techniques. DoE provides a theoretical basis for experimentation in many domains of knowledge and is particularly appropriate for studying simultaneously several variables, in order to identify the input variables with greatest effect on the output variable (the response or outcome of the experiment) and the levels at which they should be kept to improve process or product performance.

Two-level fractional factorial designs have been used as screening designs to reduce the number of input variables to a manageable few by carrying over just a fraction of all factor-level combinations. The experimental runs are often replicated to obtain an estimate of experimental error which can be used to construct statistical tests for assessing factor significance. However, when experiments are conducted in manufacturing facilities, the processes complexity often makes the replication of physical experiments prohibitive, if not impossible, due to either technical, economical or time constraints. Consequently, unreplicated factorial designs have been assuming an important role in process and product improvement.

Manuscript received February 26, 2007.
Nuno Costa, Escola Superior de Tecnologia de Setúbal, Campus do IPS, Estefanilha, 2910-761 Setúbal, Portugal (phone: 351265790000; fax: 351265721869; e-mail: ncosta@est.ips.pt).
Zulema Lopes Pereira, Faculdade de Ciências e Tecnologia - Universidade Nova de Lisboa, Monte de Caparica, Portugal (zlp@fct.unl.pt).

The identification and selection of methods for analysing unreplicated factorial designs are not simple tasks. They are scattered in many different technical and scientific journals, which may not be readily available and require financial resources to subscribe them. Fortunately, comparisons on methods performance have been reported in the literature, for example in [1]-[2], which facilitates their selection.

The main purpose of this paper is to present methods objectively selected from the literature and illustrate them through examples, so that practitioners can make objective decisions and identify the active effects in unreplicated factorial designs with a minimum number of additional runs.

II. Analysis of unreplicated factorial designs – Theoretical frame

Analysis of unreplicated factorial designs has been quite explored and still constitutes an open and active research field. The most popular tool for identifying active effects in unreplicated factorial designs is the normal or the half-normal probability plot of the contrasts. These plots have been extensively applied although several researchers recognise that they are informal and subjective. It is not easy to identify and classify as inactive the effects that fall along a straight line, and as active the ones that tend to "fall off the line". Even when all effects are inactive the plotted points do not lie perfectly on a straight line, which highlights the subjectivity in deciding what constitutes "falling off the line". Whether or not a particular contrast is included into the model may depend on practitioners' sensibility and knowledge on process and product. Thus, objective methods are preferable and numerous alternatives have been reported in the literature. It is important to avoid empirical practices and subjective analyses in experimental studies, reinforcing one important message: efficient methods whose interpretation is not subjective are more suitable and advisable, mainly for those who do not have enough background in DoE.

There are several competing methods but, in general, these methods only work well under the so-called effects sparsity principle, that is, under the assumption that only a few effects are active. This hypothesis is frequently true, but not always. In practice, there is not previous knowledge on the number and magnitude of active effects or whether abnormalities (*outliers*) exist in the data set. Furthermore, there is not one method that performs well for various configurations of number and size of active effects. For these reasons, it is advisable to apply, separately but to the same data set, methods that perform well for a small number (up to 20%) of active effects, for a large number (more than 20% up to 50%) of active effects, and also in presence of an outlier.

Misidentify an active effect as inactive (Type II error) is much more severe than misidentify an inactive effect as active (Type I error) and can compromise the efficiency of subsequent experiments. For these reasons, greater emphasis should be placed on keeping low Type II errors and accepting higher Type I errors in the screening phase. According to [2], the most powerful methods for an *EER* (Experimentwise Error Rate – error rate of declaring at least one inactive effect as active under the null hypothesis that no contrast is active) as close to 5% as possible are presented in [2]-[3].

This paper reviews and illustrates these methods and the method presented in [4], which can identify the real active effects when there are abnormalities (*outliers*) in the data. This method computes the effects based on the rank transformation of the experimental results, and then identifies the active effects applying a formal test of normality coupled with an outlier test.

To test the normality of the contrasts one should apply the statistic

$$W' = \frac{\left(\sum_{i=1}^{m} z_i c_i\right)^2}{\left(\sum_{i=1}^{m} z_i^2 \sum_{i=1}^{m} (c_i - \bar{c})^2\right)} \qquad (1)$$

where $\bar{c}$ is the average of the ordered contrasts c_i and z_i is the ith inverse normal order-statistic in a sample of size m, as defined in [5]. For a low significance level of W' test, the active effects are those whose contrasts fall outside the interval $[-2d_F, +2d_F]$ where d_F is the difference between the third and the first quartiles of c_i. The effects whose contrasts fall outside the interval are candidates for being considered active effects.

Dong [3] defined an estimator of the contrast standard error by

$$S_{Dong} = \sqrt{\frac{1}{m_{inactive}} \left(\sum_{|c_i| < 2.5 S_0} c_i^2 \right)} \qquad (2)$$

where m is the number of contrasts, $m_{inactive}$ is the number of inactive contrasts characterised by $|c_i| \le 2.5 S_0$, and $S_0 = 1.5 \times median(|c_i|)$. A contrast is declared active if

$$|c_i| > t_{\gamma; m_{inactive}} \times S_{\mathrm{Dong}} \qquad (3)$$

where $\gamma = \left(1 + 0.98^{1/m}\right)/2$.

This method is the most powerful, that is, has the greatest ability to identify all active effects and no inactive effects, which is the ideal case, for a small number of active effects, namely for one and two effects with the same or different magnitudes [2].

Chen and Kunert [2] proposed a multistage stage procedure whose test statistic $MaxUr$ is based on the generalised likelihood ratio test statistic

$$L_{m,k} = \frac{\sum_{i=m-k+1}^{m} |c_i|^2 / k}{\sum_{i=1}^{m} |c_i|^2 / (m-k)} \qquad (4)$$

where k (k = 1, 2, …, m-1) is the estimate number of active contrasts. Considering that

$$MU_k = F_{k; m-k}\left(L_{m,k}\right) \qquad (5)$$

and

$$MaxUr = \max_{1 \le k \le r}\left(MU_k\right) \qquad (6)$$

one should consider that r is the number of active contrasts if $MaxUr$ is larger than a critical value $c_{\alpha, m, r}$, where α is the significance level and m is the number of contrasts.

This method is the most powerful for a larger number of active effects, more than 3 out of 15, with the same or different magnitude [2].

III. Examples

This paper re-analyses two examples from the vast literature on DoE. The first one is an unreplicated factorial design presented in [6] and interpreted through empirical analysis of experimental results. The study aimed at identifying the effects with significant influence on the lower eutectic temperature in freeze-dried pharmaceutical products. The second one describes the application of an unreplicated factorial design in developing a nitride etch process on a single-wafer plasma etcher [7].

Example 1 – Lower Eutectic Temperature

In regard to lower eutectic temperature, Sunberg [6] concluded that B and BC were active location effects. The large magnitude of these effects is highlighted in the half-normal probability plot of contrasts (see Figure 1). In this situation, with few effects with large magnitude, Dong's and Chen and Kunert's methods also identify those effects as active. This is not surprising because there are small differences in methods performance when the active effects have large magnitudes [2]. Moreover, it is important to point out that the active effects are identified by Chen and Kunert's method for a low value of the significance level, providing much more confidence in accepting them as active effects. Aguirre and Trejo's method reveals the same conclusion (see Table 1), which was unexpected because the Rank method, in general, destroys the ability for detecting active effects when there are not outliers in the data.

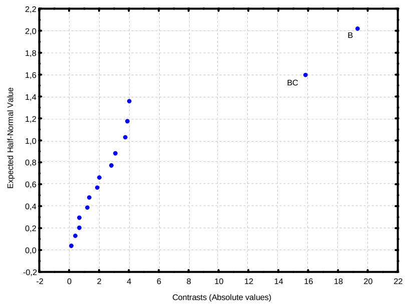

Fig. 1 – Lower eutectic temperature

Example 2 – Nitride Etch Process

In [7] is analysed the etch rate for silicone nitride in a single-wafer plasma etcher for the following variables: gas flow (A), the power applied to the cathode (B), the pressure in the reactor chamber (C), and the spacing between the anode and the cathode (D). Analysing a normal probability plot, the authors concluded that the main effects of A and D, as well as the AD interaction, were significant.

Figure 2 shows one active effect of large magnitude (D) and two active effects of smaller magnitude (A and AD). According to Table 1, the three methods lead to different conclusions. The erroneous solution of Aguirre and Trejos' method is not surprising because there are not abnormalities in the data set. The solution provided by Chen and Kunert´s method reveals five active effects while Dong´s method concludes that only three effects are active. This can be explained because, according to Chen and Kunert [2], their method may declare active some inactive effects for some configurations of active effects, which seems to be the case in this example. When Dong's and Chen and Kunert's methods present different solutions, it is necessary to run additional experiments for identifying the real active effects. Firstly, one should run an experiment at the selected levels of active effects identified by both methods and, afterwards, run the other trials by adding up each one of the remaining active effects at a time. The real active effects are identified looking at the experiments results and taking into account the objective of the study. In this example, the effects A, D and AD are identified as active by both methods. So, the first experiment should include factors A and D at their best levels, which would be identified from analysis of interaction AD. The following experiment should include factors B and C at their best levels, which would be identified from analysis of interaction BC. Since the study objective is to minimise the lower eutectic temperature, B and C are active effects if this experiment yields a value smaller than the one obtained in the previous experiment.

Table 1 – Results summary

Method	**Example 1**		**Example 2**				
Dong	B	BC	A	D	AD	-	-
Chen and Kunert	B	BC	A	D	AD	BC	ABCD
Aguirre and Trejo	B	BC	D	AD	AC	-	-

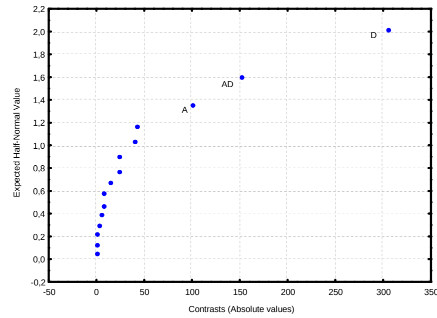

Fig. 2 – Nitride etch process

IV. Conclusions

There are configurations of active effects distinct of those ones plotted in Figures 1 and 2. Nevertheless, the authors argue that Dong's and Chen and Kunert's methods will identify all active effects for the most common configurations that arise in practice. When both methods present the same solution, this is a strong indicator of the real active effects. On the contrary, if the methods identify different active effects, the inactive effects misidentified as active can be determined through few additional trials, as illustrated in example 2. Informal procedures and graphical methods, such as normal and half-normal probability plots, allow a visual check on the accuracy of the decisions provided by other methods, being especially useful in the detection of outliers. When the normal plot of the estimated effects does not show any evident abnormality, this may mean that the data set is free of outliers and, consequently, Aguirre and Trejo's method application is unnecessary.

The methods presented in this paper may not be the magic wand to improve process and product quality. However, they are efficient, do not require any special algorithm or software to be implemented and can enhance the decision making process whenever the practitioner has to analyse the results obtained from unreplicated factorial designs.

References

[1] Hamada, M. and Balakrishnan, N. [1998]. Analyzing Unreplicated Factorial Experiments: A Review with Some New Proposals, Statistica Sinica, 8, pp. 1–41.

[2] Y. Chen and J. Kunert, "A New Quantitative Method for Analyzing Unreplicated Factorial Designs," *Biometrical Journal*, 2004, vol. 46(1), pp. 125-140.

[3] F. Dong, "On the identification of active contrasts in unreplicated fractional factorials," *Statistica Sinica*, 1993, vol. 3, pp. 209–217.

[4] V. Aguirre and M. Trejo, "Outliers and the Use of the Rank Transformation to detect active effects in Unreplicated Factorial

Experiments," *Communication Statistic-Simulation*, 2001, vol. 30(3), pp. 637-663.

[5] H. Benski, "Use of a Normality Test to Identify Significant Effects in Factorial Designs," *Journal of Quality Technology*, 1989, vol. 21, pp. 174–178.

[6] R. Sunberg, "Interpretation of unreplicated two-level factorial experiments, by examples," *Chemometrics and Intelligent Laboratory Systems*, 1994, vol. 24, pp. 1-17.

[7] D. Montgomery, and G. Runger, *Applied Statistics and Probability for Engineers*, John Wiley & Sons, ch.12.