FRACTIONAL FACTORIAL DESIGNS FOR FERTILIZER EXPERIMENTS WITH 25 TREATMENTS IN POOR SOILS

Armando CONAGIN[1]
Décio BARBIN[2]
Silvio Sandoval ZOCCHI[2]
Clarice Garcia Borges DEMÉTRIO[2]

- *ABSTRACT: In this paper, we discuss some aspects of fractional factorial designs* $5^{k-(k-2)}$*, where k is the number of factors, with only 25 treatments involving two to six quantitative factors, with the purpose of using them on experiments on poor soil areas like those of "cerrado". They are specially developed in order to assess the nutritional response to fertilizer soil addition in new areas. We also evaluate the performance of the design using simulations considering previous information.*

- *KEYWORDS: Factorial experiments; high-order interactions; alias structure; confounding; simulation; bias.*

1 Introduction

In Brazil, approximately 2 million km^2 is occupied by "cerrado" areas, its second largest biome. In those areas, the soil is an essential factor for the growth and development of vegetation and has a low concentration of nutrients nitrogen (N), phosphorus (P), potassium (K), calcium (Ca) and magnesium (Mg), and high aluminum saturation, what is a characteristic of a poor soil which tend to be acidic. As a consequence, the low nutrient concentration and high concentration of aluminum in the soil contribute to low production of biomass of plants, causing the scleromorphism of native vegetation and also influence the low production throughout the year. This means that a minimum necessary to improve production is to use fertilizers containing the macronutrients N, P, K and correct the soil acidicity by addition of some Ca source (lime). It is also important to account for the difference in production caused by different number of plants by unit of area.

With the increasing establishment of agriculture in the "cerrado", planning fertilizer experiments is of major importance. In this type of experiment, in general, at least five

[1] Instituto Agronômico de Campinas - IAC, CEP: 13012-970, Campinas, SP, Brazil.
[2] Universidade de São Paulo - USP, Escola Superior de Agricultura "Luiz de Queiroz" - ESALQ, Departamento de Ciências Exatas, CEP: 13418-900, Piracicaba, SP, Brasil. E-mail: *decio.barbin@usp.br*; *sszocchi@gmail.com; clarice.demetrio@usp.br*

quantitative factors, with three to five levels each, are used with the aim of determining the optimal economical design and studying the response surface shape.

In general there is a great deal of redundancy in a factorial experiment in that high-order interactions are likely to be negligible and some variables may not affect the response at all (Box, Hunter and Hunter, 2005, Brien, 2010a, b, Montgomery, 2012). Solutions to overcome those problems are the use of incomplete block designs and fractional factorials. Conagin and Jorge (1977, 1982a) proposed a 5^{3-1} fractional factorial to be used in fertilizer experiments and an application of it is illustrated in Caetano et al (2013). Andrade and Noletto (1986) presented $(1/2)4^3$ and $(1/4)4^4$ fractional factorials to be used in experiments to study the fertility of "cerrado" soils. Primavesi et al (2004) uses the $(1/2)4^3$ fractional factorial to design an experiment to measure the response of oats to fertilization on red yellow latosol in two planting systems.

In this paper, we discuss some aspects of fractional factorial designs $5^{k-(k-2)}$, where k is the number of factors, with only 25 treatments involving two to six quantitative factors, with the purpose of using them on experiments on poor soil areas like those of "cerrado". They are specially developed in order to assess the nutritional response to fertilizer soil addition in new areas. We also evaluate the performance of the design using simulations considering previous information.

2 Some basics

The concept of factorial experiments was introduced by Fisher (1935). While single-treatment-factor experiments involve just a single treatment factor, others involve two or more factors and are often performed as factorial experiments (Brien, 2010a, b). In those experiments the treatments are all combinations of the levels of all the factors and, generally, the number of treatments is equal to the product of the numbers of levels of the factors in the experiment. The major advantage of factorial experiments is that they allow the detection of interaction but the main disadvantage is that the total number of treatments becomes large as the number of levels and/or number of factors increases. Also, in most situations there are more factors to be investigated than can be conveniently accommodated with the time and budget available or there is an upper limit on the number of experimental units due to economical reasons or in order to have homogeneous conditions.

Full factorial experiments were, initially, proposed for two and three factors with two levels of each factor (Fisher, 1935) and extended to k factors, particularly useful in the early stages of experimental work when there are likely to be many factors to be investigated (Mongomery, 2012). Designs with 3 or more levels as, for example, 3×3, 3×4, 4×4 and 3×3×3 were proposed to study the shape of response surfaces and to estimate linear and quadratic effects and interactions. Yates (1937) presented a comprehensive survey of the simpler factorial designs and a description of the appropriate methods of analysis.

When there are four or more factors and if the experimenter can assume that certain high-order interactions are negligible, the number of treatments can be reduced by running a fraction of the complete factorial experiment. These designs are called fractional factorial designs and are among the most widely used types of designs for product and process design and for process trouble shooting (Montgomery, 2012). In the case of two-level factors, those designs are mainly used as screening experiments with the purpose of identifying those factors that have large effects. On the other hand, if the factors have three or more levels the fractional factorial designs can be used to fit response surface models (Box and Draper, 1987; Khuri and Cornell, 1996; Myers, Montgomery, Anderson-Cook, 2009).

Fractional factorial designs are expressed using the notations l^{k-p} or $\frac{1}{l^p}l^k$ or $(1/l^p)l^k$, where l^{k-p} is the number of treatments used per fraction, l is the number of levels of each factor investigated, k is the number of the factors and p describes the size of the fraction of the full factorial used. Formally, p is the number of *generators*, assignments as to which effects or interactions are *confounded*, i.e., cannot be estimated independently of each other. A design with p such generators is a $1/l^p$ fraction of the full factorial design (Box, Hunter and Hunter, 2005). However, in these designs the interactions do not go away, they just become confounded with other effects. This is not necessarily a bad thing, but it is a good idea to be aware of it so one can make an informed decision about the design one wants.

When selecting a fractional factorial design it is important to consider:

i) how many experimental units are required,

ii) which effects are aliased with effects of interest,

iii) how many effects are aliased with the effects of interest.

The best fractional factorial design is the most economical one while enabling satisfactory estimation of the effects of interest.

3 Methodology

3.1 Generating a fractional factorial design

A l^{k-p} design can be generated superimposing orthogonal Latin squares or from a full factorial structure by choosing an *alias structure* (Wu and Hamada, 2000). The use of latin squares to produce fractional factorial designs has been suggested by Cochran and Cox (1957), Davies (1950) and John (1971). This methodology was used to obtain 5^{3-1} design (Conagin and Jorge, 1977), $(1/2)4^3$ design in four blocks (Conagin and Jorge, 1982b) and $(1/2)4^3$ design in two blocks (Andrade and Noleto, 1986).

A 5^{5-3} design, for example, is 1/125 of a five level, five factor factorial design. Rather than 3125 treatments that would be required for the full factorial experiment, this experiment requires only 25 treatments. The 25 treatments can be generated superimposing three of the four orthogonal latin squares 5×5, with the addition of two columns of treatments that produces treatments with the levels of one factor balanced for the levels of the remaining factors.

An alternative method is to generate a fractional factorial design from a full factorial structure by choosing an *alias structure* that determines which effects are confounded with each other. Wu and Hamada (2000) discuss how to obtain 25-run fractional factorial designs at five levels and 49-run fractional factorial designs at seven levels, showing the results in their Tables 6C and 6D of Appendix, and generalize for l^{k-p}.

Following Wu & Hamada (2000), the 25 runs of a 5^{6-4} design are generated considering, initially, the 5×5 combinations of the levels of the first two factors, as given by the first two numbers (in bold) of the cells, presented in Table 1. The remaining four numbers of the cells are generated from the first two by using the following method. Let x_1 and x_2 denote the first two numbers, respectively. Then the third through sixth numbers, having as central point (3,3,3,3,3,3), are obtained by setting:

$$x_3 = (x_1 + 1.x_2 + 1)(mod\ 5) + 1,$$

$$x_4 = (x_1 + 2.x_2 + 3)(mod\ 5) + 1,$$

$$x_5 = (x_1 + 3.x_2 + 0)(mod\ 5) + 1$$

$$x_6 = (x_1 + 4.x_2 + 2)(mod\ 5) + 1.$$

Table 1 - A 5^{6-4} fractional factorial design with six quantitative factors and five equidistant levels (1, 2, 3, 4 e 5), giving 25 treatments, denoted by the sequence $(x_1, x_2, x_3, x_4, x_5, x_6)$

131111	**23**2222	**33**3333	**43**4444	**53**5555
142345	**24**3451	**34**4512	**44**5123	**54**1234
153524	**25**4135	**35**5241	**45**1352	**55**2413
114253	**21**5314	**31**1425	**41**2531	**51**3142
125432	**22**1543	**32**2154	**42**3215	**52**4321

This is equivalent to use a centered version of the $p = 4$ defining relations $C = AB$, $D = AB^2$, $E = AB^3$ and $F = AB^4$ or the four generators of the design $I = ABC^4 = AB^2D^4 = AB^3E^4 = AB^4F^4$ (design of resolution III) with their generalized interactions that are automatically confounded. The alias of any main effect or component of interaction is produced by the multiplication modulus 5 of the effect by I, I^2, I^3, and I^4 and using the convention that the first letter have unitary power.

Finally, to have a 5^{k-p} fractional factorial with k factors and 25 treatments, we fix $p = k - 2$ and using any k numbers that are in the same position in all the cells in Table 1 we have a $5^{k-(k-2)}$ design. Note that, for $k = 3, \ldots, 6$, we get 5^{3-1}, 5^{4-2}, 5^{5-3} and 5^{6-4} designs. Once generated the randomized level sequence, for example, $(x_6, x_3, x_4, x_5, x_2)$, we obtain the 5^{5-3} fractional factorial design.

The method suggested by Cochran and Cox (1957), Davies (1956) and John (1971), based on superimposing three of the four orthogonal latin squares with the addition of two adequate columns of treatments would use the sequences $(x_3, x_4, x_5, x_1, x_2)$, $(x_3, x_4, x_6, x_1, x_2)$, $(x_3, x_5, x_6, x_1, x_2)$ and $(x_4, x_5, x_6, x_1, x_2)$ as presented in Table 2. Note that the level sequence $(x_6, x_3, x_4, x_5, x_2)$ gives the same treatments presented in III of Table 2.

Once chosen one of the fractional factorial design, a proper randomization is needed before using it.

Table 2 - Four 5^{5-3} fractional factorial design obtained by the generated level sequences $(x_3, x_4, x_5, x_1, x_2)$, $(x_3, x_4, x_6, x_1, x_2)$, $(x_3, x_5, x_6, x_1, x_2)$ and $(x_4, x_5, x_6, x_1, x_2)$ from Table 1

I	11113	22223	33333	44443	55553
	23414	34524	45134	51244	12354
	35215	41325	52435	13545	24155
	42511	53121	14231	25341	31451
	54312	15422	21532	32142	43252
II	11113	22223	33333	44443	55553
	23511	34121	45231	51341	12451
	35412	41522	52132	13242	24352
	42314	53424	14534	25144	31254
	54215	15325	21435	32545	43155
III	11113	22223	33333	44443	55553
	24514	35124	41234	52344	13454
	32415	43525	54135	15245	21355
	45311	51421	12531	23141	34251
	53212	14322	25432	31542	42152
IV	11113	22223	33333	44443	55553
	34515	45125	51235	12345	23455
	52414	13524	24134	35244	41354
	25312	31422	42532	53142	14252
	43211	54321	15431	21541	32151

3.2 Data analysis

To analyze the data from a fractional factorial design we consider the classical linear model

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{e},$$

where $\boldsymbol{y}$ is the vector of observations, $\boldsymbol{X}$ is the design matrix, $\boldsymbol{\beta}$ is the parameter vector and $\boldsymbol{e}$ is the error vector, normally distributed with mean $\boldsymbol{0}$ and variance-covariance matrix $\sigma^2\boldsymbol{I}$, $\boldsymbol{e} \sim \mathrm{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I})$. The vector of estimated parameters is given by

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$$

with $\mathrm{E}(\hat{\boldsymbol{\beta}}) = \mathbf{0}$ and $\mathrm{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2\boldsymbol{I}$. The analysis of variance can be done in the usual way (Steel and Torrie, 1981, Draper and Smith, 1966, Montgomery, 2012) using statistical packages as R and SAS, for example.

For a quantitative factor, X, in a five-level design, its linear and quadratic effects can be represented by the orthogonal contrast vectors $X_1 = \frac{1}{\sqrt{10}}(-2,-1,0,1,2)$ and $X_2 = \frac{1}{\sqrt{14}}(2,-1,-2,-1,2)$. Then, for example, the quadratic regression model without interaction for the fractional factorial 5^{5-3} can be expressed as

$$E(Y_i) = \beta_0 + \beta_{11}X_{11i} + \cdots + \beta_{15}X_{15i} + \beta_{21}X_{21i}^2 + \cdots + \beta_{25}X_{25i}^2 \quad (i = 1,\ldots,25),$$

where β_0 is the intercept, β_{1k} and β_{2k}, $k = 1,\ldots,5$, are, respectively, the linear and quadratic parameters for the k^{th} factor, X_{1ki} and X_{2ki} are, respectively, the values of the linear and quadratic polynomials. Then the sums of squares for a linear and quadratic regressions can be obtained, respectively, by

$$SSLR_k = \frac{\left(\sum_{i=1}^{25} X_{1ki}Y_i\right)^2}{\sum_{i=1}^{25} X_{1ki}^2} \; and \; SSQR_k = \frac{\left(\sum_{i=1}^{25} X_{2ki}Y_i\right)^2}{\sum_{i=1}^{25} X_{2ki}^2}, \; (k = 1,\ldots,5).$$

This means that only 2 degrees of freedom (df) out of 4 df of any main effect are used, giving a total of $(2 * k)$ df for the model and leaving $2 * (12 - k)$ df for the error under the assumption of no interaction.

3.3 Simulation study

In order to evaluate the performance of fractional factorial 5^{5-3} designs, a Monte Carlo simulation study of size equal to 10,000 was performed. Using information from previous corn experiments, the common factor levels for this type of experiments are: dosages for N (30, 45, 60, 75, 90) kg.ha^{-1}, P_2O_5 (30, 45, 60, 75, 90) kg.ha^{-1}, K_2O (30, 40, 50, 60, 70) kg.ha^{-1}, lime (1;1,5; 2,0; 2,5; 3) t.ha^{-1} and population density (50, 55, 60, 65, 70) $\times 10^3$ plantas.ha^{-1} with the vector of parameters equal to

$$\beta = [5952{,}0000 \;\; 975{,}8074 \;\; 763{,}6753 \;\; 827{,}3149 \;\; 763{,}6753 \;\; 1081{,}8734 - 1003{,}9920 \\ - 896{,}4215 \;\; - 376{,}4970 \;\; - 322{,}7117 \;\; - 376{,}4970]^T,$$

and the error $e_i \sim N(0;\, \beta_0 \times \mathrm{CV})$ for $i = 1,\ldots,25$, that is, $e_i \sim N(0; 5952 \times CV)$ for the coefficients of variation, CV, assuming the values 0,1%, 0,2%, ... ,15,0%.

For each combination of simulation factors, we calculated:

i) the mean absolute bias,

ii) the percentage of times that the signs of the parameter estimates for the quadratic terms, $\boldsymbol{\beta}_2$, were negative,

iii) the percentages of rejection of the null hypothesis H_0: $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \mathbf{0}$, considering the significance levels 5%, 10% e 20%, of the estimates over the 10,000 samples.

4 Results

The results of the simulation study are summarized in Figure 1, showing that as the coefficient of variation increases:

i) the mean absolute bias percentage of parameter estimates increases,

ii) the percentage of times that the signs of the parameter estimates for the quadratic terms, $\beta_{21}, \ldots, \beta_{25}$, were negative decreases,

iii) the percentage of rejection of the null hypothesis $H_0: \beta_{11} = \cdots = \beta_{15} = \beta_{21} = \cdots = \beta_{25} = 0$, considering 5%, 10% and 20% significance levels, decreases.

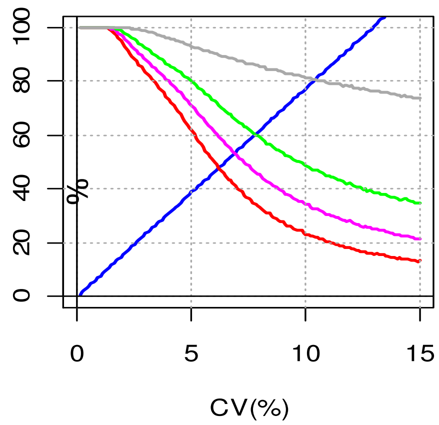

Figure 1 - Mean absolute bias percentage of parameter estimates (---), percentage of times that the sign of the parameter estimates for the quadratic terms, $\beta_{21}, \ldots, \beta_{25}$, were negative (---) and percentages of rejection of the null hypothesis $H_0: \beta_{11} = \cdots = \beta_{15} = \beta_{21} = \cdots = \beta_{25} = 0$, considering 5% (---), 10% (---) and 20% (---) significance levels, as a function of the coefficient of variation (CV). Each point of the graphics was based on a 10,000 size simulation.

Then, for example, for a 5% coefficient of variation we can have around 40% absolute bias percentage of parameter estimates and 70% of rejection of the null hypothesis $H_0: \beta_{11} = \cdots = \beta_{15} = \beta_{21} = \cdots = \beta_{25} = 0$, for the 10% significance level.

From these results it is important to note that when the coefficient of variation is small, although there is bias on the estimation of the linear and quadratic regression parameters, due to confounding with some two-factor and higher interactions, there is a good chance of having negative sign for the parameter estimates of the quadratic terms, $\beta_{21}, \dots, \beta_{25}$ what is an indication of a maximum for the surface. The biggest advantage of this type of experiment, in general used as a preliminary trial, is that it uses only 25 treatments instead of 3125 treatments that would be required for the full factorial experiment.

5 Concluding remarks

The proposed design may be ideal to use to choose the dosages of fertilizers to be applied in poor soils like the "cerrado" ones where the poorness of the nutrients is a permanent situation. It could also be adopted for experiments in physiology, forestry and horticulture, or with perennial crops, as well as in physic-chemical complex ones, provided that interactions do not exist or have little influence on the responses. In experiments with fertilizers the main objective is to get the maximum production and the economical analysis.

For a study of four factors like N, P, K and Ca, the 5^{4-2} design is the appropriate choice. In the case of more factors, for example, to study additionally the importance of the population size of individuals the factors may be N, P, K, Ca and population density, the 5^{5-3} must be chosen.

The use of the 5^{5-3} design with only 25 treatments, instead of 3125 for five factors with five levels each, is subject to have bias in the estimation of linear and quadratic regression coefficients due to confounding with two-factor and higher interactions, but has a good chance of having negative sign for the parameter estimates of the quadratic terms, what is an indication of a maximum, for a low coefficient of variation.

If the levels of nutrients are chosen in a suitable range it is normally true that the crops will develop favorably with no or very small two factor interaction. In this favorable situation it is reasonable to suppose that there is no interaction in the model and the important points to consider are to estimate the value of the regressions to get a practical decision in the analysis viewing the maximum and the best economical dosage of the fertilizer utilized.

To get designs with low coefficient of variation it is always important to care about time of planting, soil preparation, adequate choice of factor dosages and excellent management of the field experiments. In greenhouse experiments, for which the water supply is adequate, the experiments exhibit always, low coefficient of variation and then the results will produce more accurate regression estimates; if samples of soil are used, physiological responses to each soil will be of great value for a posterior choice of the places to be included in a set of field experiments with fertilizers.

The performance of this type of experiment could be improved by the use of replicates of the central point (33333) and using different fractions of the full factorial in different experiments.

CONAGIN, A.; BARBIN, D.; ZOCCHI, S. S.; DEMÉTRIO, C. G. B. Fatoriais fracionados com 25 tratamentos em solos pobres. *Rev. Bras. Biom.*, São Paulo, v.32, n.2, p.180-189, 2014.

▪ RESUMO: Neste artigo, são discutidos alguns aspectos dos delineamentos fatoriais fracionados $5^{k-(k-2)}$, em que k é o número de fatores, com somente 25 tratamentos envolvendo de dois a seis fatores, com o propósito de usá-los em experimentos em solos pobres, como aqueles do cerrado. Eles são, especialmente, desenvolvidos a fim de acessar a resposta nutricional da adição de fertilizante no solo. É avaliada, também, a performance do delineamento, usando simulações, considerando informação prévia.

▪ PALAVRAS-CHAVES: Fatoriais fracionados; interações de alta ordem, estrutura “alias”; confundimento, simulação; viés.

References

ANDRADE, D. F.; NOLETO, A. Exemplo de fatoriais fracionados $1/2(4^3)$. Ajuste de modelos polinomiais quadráticos. *Pesquisa Agropecuária Brasileira*, Brasília, v.21, n.6, p.677-682, 1986.

BOX, G. E. P.; DRAPER, N. R. *Empirical Model-Building and Response Surfaces*. New York: John Wiley & Sons, 1987. 669p.

BOX, G. E. P.; HUNTER, J. S., HUNTER, W. G. (2005) *Statistics for Experimenters: Design, Innovation, and Discovery*. New Jersey: John Wiley & Sons, 2005. 639p.

BRIEN, C. J. Factorial experiments. URL: http://chris.brien.name/ee2. Accessed Dec. 2013, 2010a.

BRIEN, C. J. Factorial designs at two levels. URL: http://chris.brien.name/ee2. Access: dec. 2013, 2010b.

CAETANO, L. C. S.; VENTURA, J. A.; COSTA, A. F. S.; GUARÇONI, R. C. Efeito da adubação com nitrogênio, fósforo e potássio no desenvolvimento, na produção e na qualidade de frutos do abacaxi 'Vitória'. *Revista Brasileira de Fruticultura*, v.35, n.3 p.1-5, 2013.

COCHRAN, W. G.; COX, G. M. *Experimental design*. 2ed. New York: John Wiley & Sons, 1957. 611p.

CONAGIN, A.; JORGE, J. P. N. Delineamentos 1/5(53). *Bragantia*, v.36, n1,p.1-4, 1977.

Conagin, A.; Jorge, J. P. N. Delineamentos (1/5)(5x5x5) em blocos. *Bragantia*, v.41,n.1, p.1-3, 1982a.

CONAGIN, A.; JORGE, J. P. N. Delineamentos 1/2(43) em blocos de 8 unidades. Instituto Agronômico, Campinas, Boletim Científico n36, 1982b.

DAVIES, O. L. *Design and analysis of industrial experiments*. 2nd ed. New York: Hafner Publishing Company, 1956. 636p.

DRAPER, N. R.; SMITH, H. *Applied Regression Analysis*, New York: John Wiley & Sons, 1966. 407p.

FISHER, R. A. The *Design of Experiments*. 2ed. London: Oliver and Boyd, 1935. 260p.

JOHN, P. W. M. *Statistical Design and Analysis of Experiments*. New York: Macmilan, 1971. 356p.

KHURI, A. I.; CORNELL, J. A. *Response Surfaces*. 2ed. New York: Marcel Dekker, 1996. 510p.

MONTGOMERY, D. C. *Design and Analysis of Experiments*. 8ed. New York: John Wiley & Sons, 2012. 730p.

MYERS, R. H.; MONTGOMERY, D. C.; ANDERSON-COOK, C. M. *Response Surface Methodology. Process and Product Optimization Using Designed Experimen*t. 3ed. New Jersey: John Wiley & Sons, 2009. 680p.

PRIMAVESI, A. C.; PRIMAVESI, O.; CANTARELLA, H.; GODOY, R. Resposta da aveia branca à adubação em latossolo vermelho-amarelo em dois sistemas de plantio. *Revista Brasileira de Zootecnia*. v.33, n.1, p.1-10, 2004.

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.

SAS Institute Inc. System of Microsoft Windows, release 8, Cary, N.C. USA, 1999.

STEEL, R.; TORRIE, J. (1981) *Principles and Procedures of Statistics*. New York: Mc Grow Hill Book Company, 1981. 481p.

YATES, F. Design and Analysis of Factorial Experiments. *Tech. Comm*. No. 35, London: Imperial Bureau of Soil Science, 1937.

WU, C. F. J.; HAMADA, M. *Experiments. Planning Analysis and Parameter Design Optimization*. John Wiley & Sons, 2000. 630p.

Received in 05.02.2014

Approved after revised in 17.06.2014