

Andrzej Skowronek (skowronek@mech.pk.edu.pl)
Faculty of Mechanical Engineering, Cracow University of Technology

APPLICATION OF THE MULTI-GENERATING METHOD
TO SUPPORT THE QUALITY OF SMART DESIGNS OF EXPERIMENT

ZASTOSOWANIE METODY WIELOKROKOWEGO GENEROWANIA
WSPOMAGAJĄCEJ JAKOŚĆ ELASTYCZNYCH PLANÓW EKSPERYMENTU

Abstract

The article presents a method used in the process of generating smart designs of experiment in a dedicated computer program with the application of pseudo-random numbers, the use of which may have a considerable impact on the quality of generated designs. In order to increase the probability of generating the optimal design for the defined parameters of generating, a special method was applied, in which several designs were generated and the best of them selected, based on the equipartitional analysis parameters. The results confirm a significant positive influence of the analyzed method on the smart design's quality.

Keywords: smart design of experiment, experimental research

Streszczenie

Artykuł przedstawia metodę używaną w procesie generowania elastycznych planów eksperymentu w specjalnym programie komputerowym z zastosowaniem liczb pseudolosowych, których użycie może mieć znaczący wpływ na jakość planów. W celu zwiększenia prawdopodobieństwa wygenerowania planu optymalnego dla ustalonych parametrów w generowaniu planu, zastosowano specjalną metodę, w której generowanych jest wiele planów i wybierany najlepszy na podstawie parametrów tzw. analizy ekwipartycyjnej. Uzyskane wyniki potwierdzają znaczący pozytywny wpływ metody na jakość elastycznych planów eksperymentu.

Słowa kluczowe: elastyczne plany eksperymentu, badania eksperymentalne

1. Introduction

Experimental research is one of the commonly used source of obtaining information. Special techniques, known as design of experiment methodology (DoE), are applied to conduct experimental research. Planning of experimental research is based on the application of special designs of experiment, which allow for a significant reduction in size of experiment (runs of experiment, observations, number of measurements, etc.). It may be particularly important in the case where a researcher is not able to perform the research for all combinations of input levels, which could be, for example, a result of restrictions imposed on the time of experiment's realization, its high costs or just inability to realize certain combinations of input factors' levels. It should be noted that reducing the size of the experiment does not necessarily lead to the reduction in the amount of information obtained as a result of the experiment conducted with the application of DoE techniques.

Depending on the goal of experimental research, various types of designs can be used [1–3]. When using the traditional design of the experiment, the researcher must accept its characteristics and execute the experiment strictly according to the design which has been selected. In particular, such design's characteristics as the number of units, the input factor levels and their number cannot be changed, as this could make the experiment difficult or even impossible to realize. Quite a different approach to the conception of experiment planning is applied when smart designs of experiments are used [4]. The researcher is allowed to set the number of design's units and the number of input levels. Moreover, one can impose restrictions on the input space and avoid combinations of factor levels which are not allowed or not feasible to be implemented. The main idea of smart designs of experiment is the possibility of easy application in experimental research. The researcher only needs to define the most important features of the experiment, such as the number of runs, the number of input factors, levels of input factors and possibly some restrictions put on the input space.

2. The Idea of Smart Designs of Experiments

Smart designs of experiment are generated in a dedicated computer application, based on three important principles: adaptation, randomness and equipartition [4, 5]. The first principle means the possibility of adjusting the design's characteristics to the conditions of the experiment and characteristics of the analyzed object. The researcher is able, for example, to set the number of design's units and the number of its levels for each input. The second principle means that smart designs are created in a non-deterministic manner: both the generation of input levels and the selection of design's units are conducted with the use of pseudo-random numbers. However, there are some limitations put on the random way of generation of design's units:

- ▶ a parameter called “important difference” (Δx), a minimal permissible distance between the last generated value and existing values of each input factor levels,

- ▶ a parameter called “minimal Euclid’s distance” ($esmin$) – it is Euclid’s distance to the nearest “neighbor-unit” in the input space, calculated for each design’s unit, each unit must fulfill the $esmin$ condition: $es \geq esmin$

The conceptions of both parameters described above are based on the Euclid’s distance measure and they use the fact that a set of experimental design units in the input space is equivalent to the set of points in the orthogonal coordinate system as well as the combinations of input levels (which make up the units of designs) are equivalent to the points coordinates. The Δx and $esmin$ parameters support equipartition of the design’s units in the input space. If there are no other assumptions, design’s units should cover regularly the whole input space (the third rule). The equipartition of design’s units means regularity and uniformity of the design’s units in the input space, which reduces the likelihood of occurrence of empty spaces (without any design’s unit), and is strongly required if you want to obtain as much data on the research object as possible and find out the research object function which specifies how inputs affect the output. The equipartition of units is the main smart design optimality criterion.

To estimate the regularity of the distribution of the design’s units (equipartition), the method of equipartitional analysis (EPA) is used [4, 5]. This analysis provides parameters that allow you to make a qualitative assessment of the generated designs. The high equipartition of the units distribution in the input space means the high quality of the design. In the equipartitional analysis, the created design of experiment is compared to the master-design, whose units are distributed perfectly regularly in the input space. The master-design always has the same number of inputs as the analyzed designs and the same number of input factors’ levels, but the number of master design’s units is usually significantly higher and equal to the product of numbers of all input levels (number of all combinations of input factors’ levels). For example, in case of design consisting of 10 units, 6 levels for the first factor and 8 for the second factor, the master design consists of 48 units (all combinations of 6 factors’ levels for the first input factor and 8 for the second factor). The input factors’ levels of the master-design are calculated for each input factor by regular division of the length of input factors’ ranges and the number of factors’ levels (Fig. 1).

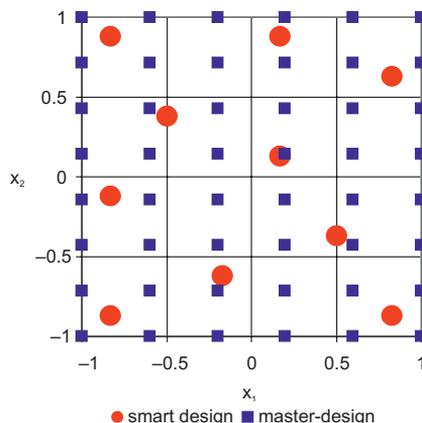


Fig. 1. Smart design and master design

For each unit of the master-design (x^m), one can evaluate the Euclid's distances to all units of the analyzed smart design (x^s):

$$e(x^m, x^s) = \sqrt{\sum_{i=1}^d (x_i^m - x_i^s)^2} \quad (1)$$

where:

d – number of input factors.

Having calculated matrix E of Euclid's distances between master design units and smart design units:

$$E = \begin{bmatrix} e(x_1^m, x_1^s) & e(x_1^m, x_2^s) & \dots & e(x_1^m, x_{ns}^s) \\ e(x_2^m, x_1^s) & \dots & & \\ \dots & & & \\ e(x_{nm}^m, x_1^s) & e(x_{nm}^m, x_2^s) & \dots & e(x_{nm}^m, x_{ns}^s) \end{bmatrix} \quad (2)$$

where:

nm – number of master design units,

ns – number of smart design units,

one can evaluate vector $E1$ of the Euclid's distances between all master design units and their nearest smart design unit:

$$E1 = [\min(e(x_1^m, x_1^s), \dots, e(x_1^m, x_{ns}^s)), \dots, \min(e(x_{nm}^m, x_1^s), \dots, e(x_{nm}^m, x_{ns}^s))] \quad (3)$$

For such a collection ($E1$ vector, called equipartitional set), one can evaluate a lot of statistical parameters, e.g. descriptive statistics [6] or make one of the statistical tests [7], which could be an equipartition criterion in this analysis. Two parameters have been used: the maximal ($e1max$) and mean ($e1mean$) value of the equipartitional set. The $e1mean$ parameter describes the central tendency of the equipartitional set whereas the $e1max$ parameter provides the information whether there are any huge empty areas in the input space (without any design's units), which is important taking into consideration the assumption that the design's units should regularly cover the whole input space. The dependence between both parameters and the design's quality (quality means equipartition, perfect regularity of the design's units distribution in the input space) was verified and described in [5]. The conclusion was that the less value of equipartitional parameters means the more regular distribution of design's units in the input space. Assuming that smart designs quality is identified with regularity and uniformity of design's units distribution in the input space (equipartition), the quality can be evaluated using equipartitional analysis parameters, for example $e1max$ and $e1mean$, which should be minimized. Both parameters can be used separately and each could be the main criterion of design quality. However, it is recommended to use them together.

There are three methods of generating the inputs' levels used to create smart design's units in the current version of designs' generator. In the Z-method, inputs' levels are generated as pseudo-random values from the normalized range $[-1, 1]$ and checked if they pass the important difference condition test. If a value (factor's level) fails the test, it is removed and the next one is generated to reach the right amount (assumed number of factor levels). In

the “R” method, the input factors’ levels are calculated by dividing the input ranges by the demanded numbers of input factors’ levels. The first level is calculated as the minimum of the input range, whereas the last level is calculated as the maximum of the input range. In the R2-method, the idea of levels calculation is that each level should be the center point of equal areas of influence. The first and the last levels are not equal to the minimum or maximum of the input factor’s range.

The smart design’s generator in the current version has implemented functionalities which support the selection of the optimal values of important generation’s parameters: the important difference (Δx , used in Z-method of levels’ generating) and the minimal Euclid’s distance (*esmin*), used to enhance high regularity and equipartition of design’s units in the input space [8]. When using the previous versions of the generator, a researcher must set Δx and *esmin* parameters by himself, which could make the generated design not optimal – designs’ units do not cover equally the whole input space in the case of setting to small values of generation parameters, or it is not possible to obtain the design with the assumed properties (number of units) otherwise. In the current version, the problem has been fixed since the initial values of both parameters (Δx and *esmin*) are calculated automatically.

3. A procedure of multi-generation of the smart design of experiment

Smart designs of experiment are generated in a special computer program with the application of pseudo-random numbers, the use of which may have a considerable impact on the quality of the generated designs. Computer-generated pseudo-random numbers, which have properties similar to random numbers, are deterministic. The choice of a seed value, which is used to initialize a pseudo-random number generator, seems to be crucial. A generated sequence of pseudo-random numbers can be reproduced if the used seed value is not changed [9], which should be considered an advantage. But on the other hand, randomly generated series of numbers used in the procedure of smart designs generating can cause a significant change in potential possibilities of generating a design. It may result in obtaining a better or worse quality design for the same settings of generation, which of course is a disadvantage.

In the smart designs’ generator, pseudo-random numbers are applied in the module of generating the input factors’ levels using the Z-method and in the module of unit selection from the set of candidate units (a set of units created as all combinations of all factors’ levels). Designs generated with the same seed of a pseudo-random number generator, the same parameters of generation (values of Δx and *esmin*, method of input’s levels generating) and the same design’s characteristic (the number of inputs, the number of input’s levels, the number of design’s units and method of levels generating) will be identical. However, in the case of using a different seed value, it is possible (but rather certain) to obtain various designs for the same settings defined for the generating process. Moreover, there is no certainty whether designs obtained for defined parameters of generating are really optimal (best quality) and the difference in their quality could be sometimes significant. Considering the above, it seems to be necessary to generate several designs and select one, based on EPA-parameters (*e1max* and *e1mean*), which are the



measures of designs' quality. That is the idea of multi-generating smart designs of experiment. Thus, the role of a multi-generating procedure is to eliminate or at least reduce the impact of the use of pseudo-random numbers on the generated designs' quality.

Users of smart designs generator can set a pseudo-random numbers generator seed value by themselves or can let it generate automatically, based on the real-time clock, which is a default and recommended option. To regenerate the identical design again, only its seed value must be known. The researcher can select the EPAparameter which he prefers to identify the best design. In the current version of smart designs generator, the procedure of multi-generating consist of up to 20 design generating attempts to get 10 candidate-designs using various seed values. The initial seed value is generated using system time and increases by 1 for each attempt of design generation. For each design, EPA-parameters ($e1max$, $e1mean$) are calculated and saved. After completing 10 candidate-designs (or after completing 20 attempts of generating) the best one is selected based on EPA-parameters. The one with the lowest values of $e1max$ and $e1mean$ parameters is saved as a result of the multi-generation procedure. However, theoretically it is possible to obtain no design or only one or two designs. In such cases, it is recommended to repeat the generating procedure once again.

4. Computer simulation to evaluate the effectiveness of the considered method

In order to evaluate the effectiveness of the analyzed method and its influence on the quality of generated designs, a computer simulation was performed. In the simulation, 108 smart designs were generated, using the Z, R and R2-methods of generating the input factors' levels, various numbers of factors' levels and various numbers of units (see Table 1). For each combination of four mentioned design's characteristics (number of factors, number of factors' levels, number of units, method of levels generating) a generation process was executed 2 times to check if the obtained results are both stable and repeatable.

Table 1. Combinations of design's characteristics applied in simulation

Number of input factors	Number of factors' levels	Number of designs' units	Methods of levels' generating
2	5, 5	10, 15, 20	Z, R, R2
2	6, 8	10, 15, 20	Z, R, R2
3	5, 5, 5	10, 30, 50	Z, R, R2
3	6, 7, 8	10, 30, 50	Z, R, R2
4	5, 5, 5, 5	100, 200, 300	Z, R, R2
4	5, 6, 7, 8	100, 200, 300	Z, R, R2

Despite 20 attempts to generate 10 designs, it can happen for certain combinations of designs' characteristics and initial seed values that less than 10 designs fulfilling all assumptions are generated. It was assumed in the simulation that if the number of designs achieved in the

procedure of multi-generating is less than 3, it was repeated, but only up to 2 times. If during 3 runs less than 3 designs were generated for some combination of design's characteristics, no further action was taken.

The aim of application of multi-generation procedure is supporting smart designs' quality, which is evaluated based on two EPA-parameters: $e1max$ and $e1mean$. This is the reason why both EPA-parameters were used to assess the effectiveness of the analyzed method. The effectiveness was calculated in percentage as the relative difference between values of EPA-parameters obtained for the design qualified as the worst one (the highest values of both EPA-parameters) and values of EPA-parameters obtained for the design qualified as the best one (the lowest values of $e1max$ and $e1mean$ parameters) from a set of multi-generated candidate designs (up to 20 attempts of generating to get 10 designs) for each combination of 4 designs' characteristics. In case when the value of $e1max$ parameter for some design is higher and the value of $e1mean$ parameter is lower than for the compared design in candidate-design set, the higher value of $e1max$ EPA-parameter has determined the worse design. For both sets of relative differences, (for $e1max$ and $e1mean$ EPA-parameters) two basic descriptive statistics were calculated: maximal value of effectiveness ($effmax$) and average value of effectiveness ($effavg$).

5. Results of the simulation

The summary of simulation results according to various criteria are shown in Table 2 and Table 3.

Table 2. Results of simulation obtained in the first and second run of experiment

Runs	Method's effectiveness evaluated using EPA-parameters in [%]			
	$e1max$		$e1mean$	
	$effmax$	$effavg$	$effmax$	$effavg$
1. run	39	17	18	4
2. run	43	20	26	4
overall	43	19	26	4

The effectiveness calculated in simulation for all 108 cases reached on average 19% for $e1max$ parameter and 4% for $e1mean$ parameter, whereas the maximal effectiveness for $e1max$ parameter reached 43% and 26% for $e1mean$ parameter. It is important, because decreasing EPA-parameters means improving the design's quality.

As it was described above, every generating process (for all combinations of designs' attributes) was executed twice. Let us now focus on Table 2, which shows the results obtained in the first run, second run and overall. The values of improvements obtained in both runs are similar, which confirms that the effects of application of the analyzed method are stable and repeatable.

Table 3. Results of simulation depend on the method of levels' generation

Method of levels generating	Method's effectiveness evaluated using EPA-parameters in [%]			
	e1max		e1mean	
	effmax	effavg	effmax	effavg
Z	41	18	26	7
R	43	17	12	2
R2	43	20	13	3

Comparing the results obtained for all 3 various methods of generating the factors' levels (Table 3), one can observe similar effectiveness values, especially those achieved for *e1max* parameter. For *e1max* parameter, the R2-method of generating produces the highest effectiveness, whereas for *e1mean* parameter the best results are obtained in case of the Z-method. However, you must remember that the levels of the factors in the Z-method are re-generated on each attempt of smart design creating, as opposed to the methods of R and R2, where the generated sets of factor levels are the same for the whole procedure of designs' multi-generating (up to 20 attempts, if necessary) to select the best one when the multi-generation method is active. The re-generating of levels in the Z-method can cause high variability of the results and high values of effectiveness achieved for *e1mean* parameter.

Table 4 shows relative frequencies of effectiveness cases calculated for both EPA-parameters. For *e1mean* parameter, cases with effectiveness of less than 5% are dominating, whereas for *e1max* parameter cases with effectiveness between 10% and 30% are dominating. As can be noticed, larger values of effectiveness were obtained for the *e1max* parameter, but considering the manner of calculating both statistics it appears foreseeable and justified.

Table 4. Relative frequencies of effectiveness cases for EPA-parameters

EPA-parameter	Relative frequencies of effectiveness cases for EPA-parameters (in %)				
<i>e1max</i>	range	≤ 10%	≤ 30%	≤ 50%	> 50%
	frequency	21	64	15	0
<i>e1mean</i>	range	≤ 5%	≤ 10%	≤ 15%	> 15%
	frequency	74	14	7	5

If only one design for some combination of design's characteristics and initial seed value was generated (3 of 108 cases in the simulation), the effectiveness is 0%. However, it does not mean design's poor quality. The procedure of generating smart designs of experiment includes various tools supporting quality of design, e.g. important difference condition, minimal Euclid's distance condition, automatic selection of important difference and minimal Euclid's distance values (see chapter 2). Comparing these designs to designs generated on the other run or with the application of the other method of factor levels generating, there were usually no significant differences of both EPA-parameters. It suggests a tendency that generating

only 1 design during 20 attempts provides usually some of good quality, notwithstanding the fact that multi-generating procedure does not work. In cases where several designs were obtained in up to 20 attempts of generating, a larger variety of results is usually observed. In such cases, the procedure of multi-generating works very well. There are also 9 cases where only 2 designs were generated. However, in 8 cases from these 9, some effectiveness of at least one of two EPA-parameters was observed. Generally, in 88% of cases, some effectiveness of eI_{max} parameter was observed, and in 64% of cases some effectiveness was observed for both EPA-parameters.

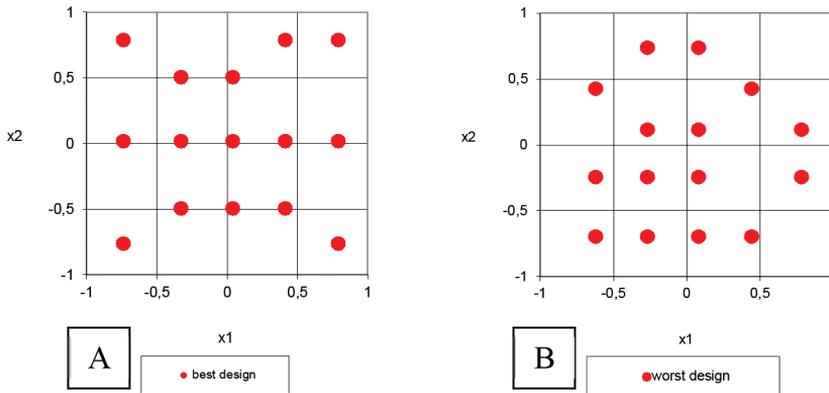


Fig. 2. Design qualified as the best of series (A) and as the worst of series (B) in multi-generation method

Figure 2 shows the distribution of the design's units in the 2-dimensional input space for a smart design generated with the following assumptions: 2 input factors, 15 units, 5 levels for each factor, Z-method of factor levels generating. The best design was selected from among 5 designs, the achieved effectiveness of multi-generating method amounted to 36% for eI_{max} and 26% for eI_{mean} , which are one of the highest values in case of eI_{max} parameter and the top value in case of eI_{mean} parameter. The design qualified as the best of series generated with the application of multi-generation method is shown in Figure 2-A, whereas the design qualified as the worst in the same series is shown in Figure 2-B. It seems that the points representing design' units are distributed really more equally (uniformly) in Figure 2-A than in Figure 2-B, which confirms the effectiveness of the method analyzed.

6. Conclusions

The results of simulation have confirmed a significant impact of the multi-generation method on enhancing smart designs of experiment's quality. The application of the considered method allows you to select the optimal design among several generated for the defined design's characteristics and, as a result, to avoid accidental impact of random numbers, and thus supports the increase in the quality of smart designs of experiment.

Considering eI_{max} parameter, the decrease of its value is usually very high and reaches more than 10% in 79% of cases in the simulation. Considering eI_{mean} parameter, the decrease is not too huge but significant as well. Significant values of effectiveness obtained for eI_{mean} parameter can suggest reducing empty areas in the input space as a result of the application of analyzed multi-generating procedure.

When it is impossible to generate more than one design in 20 attempts, the only generated design is usually of good quality, notwithstanding the fact that the multi-generating procedure does not work in such a case. But if several designs are obtained in up to 20 attempts of a multi-generating procedure, a high variety of EPA-parameters is usually observed. In such cases, the procedure of multi-generating works very well and leads to enhancing the quality of smart designs.

You should also remember that the occurrence of empty zones in the input space may be the reason why the designs are generated with values of parameters set by the researcher, in particular: a defined number of units, defined number of input factor levels and their specified values, which could in fact more or less promote uniformity of design's units in the input space and could have a negative impact on the efficiency of the analyzed method. We do not know exactly of what quality would be a smart design generated without the application of the multi-generating procedure, maybe quite good, maybe very high, but certainly the application of the multi-generating procedure cannot cause a decrease in the design's quality. The likelihood of an increase in design's quality is significant and the damaging influence of pseudo-random numbers is seriously reduced.

References

- [1] Montgomery D.C., *Design and Analysis of Experiments*, John Wiley & Sons, Hoboken 2012.
- [2] Hinkelmann K., Kempthorne O., *Design and Analysis of Experiments. Volume 1. Introduction to Experimental Design*, John Wiley & Sons, Hoboken 2008.
- [3] Hinkelmann K., Kempthorne O., *Design and Analysis of Experiments. Volume 2. Advanced Experimental Design*, John Wiley & Sons, Hoboken 2005.
- [4] Polański Z., *Empirical research – methodology and computer aiding*, [in:] *Modern metrology*, J. Barzykowski (ed.), WNT (in Polish, Warszawa 2004).
- [5] Skowronek A., *Optymalizacja procesu generowania elastycznych planów eksperymentu (Optimization the process of generating the flexible designs of experiment)*, *Czasopismo Techniczne*, 1-I/2007, 63–74 (in Polish).
- [6] Johnson R.A., Bhattacharyya G.K., *Statistics: Principles and Methods*, John Wiley & Sons, Hoboken 2010.
- [7] Sheskin D.J., *Handbook of parametric and nonparametric statistical procedures. 3rd ed.*, Chapman&Hall/CRC, Florida 2004.
- [8] Skowronek A., *A method of self-adaptation calculating the value of the smart designs of experiment generation's parameter*, *Polish Journal of Environmental Studies*, Vol. 18, No. 3B, 2009, 327–331.
- [9] Dixit J.B., *Solutions to Programming in C and Numerical Analysis*, Laxmi Publications, 2006.