2^k Factorial Design Used to Optimize the Linear Mathematical Model through Active Experiment

Ioan MILOSAN
Transilvania University of Brasov, Romania, milosan@unitbv.ro

Abstract
The data presented in this paper is part of a comprehensive study on mathematical modelling by active experiment, using first-order designs in order to optimize linear mathematical models. In this respect, a 2^k full factorial design was used. In the research, the experimentation plan was established to enable a higher number of experiments than the number of coefficients to be determined. There was established the matrix of the designed experiment and the calculation was performed according to the methodology specific to the 2^3 full factorial experiments. The results obtained by experimental design were employed in the study, using a special Mo-Ni alloyed cast iron, heat treated, whose elongation values were determined. Finally, a small number of determinations were used to determine the specific technological parameters and the maximum hardness of the analysed material.

Keywords
DOE, 2^k factorial design, cast iron, elongation

1. Introduction

Where, as part of the scientific research, statistical methods are employed in all stages of an experiment (before, during, and after the experiment), the work is carried out according to the following sequence:

- determination of the number of experiments and the conditions for their conduct (prior to the experiment);
- processing of the results (during the experiments);
- determination of the conclusions on the execution of future experiments (after the experiment).

This specific manner of conducting research is called active experiment and involves the design of the experiment conducted by [1-9]:

- establishing the necessary and sufficient number of experiments and the conditions for their conduct;
- determination by statistical methods of the regression equation, representing a certain degree of approximation, computable, the model of the process;
- determination of the conditions for obtaining the optimal performance for the analysed process.

The design of experiment (DOE) will be used in this paper with 2^k full factorial experiments (FFE) to determine the direction of movement to an optimal range from a known point, calculation performed in the research regarding the optimization of the elongation values of a heat treated special cast iron.

2.Research Objectives

The main objective of this paper is to achieve mathematical modelling by active experiment, using 2^3 full factorial experiments, in order to optimize the linear mathematical model, determining the direction of movement to an optimal domain from a known point. In this case, of interest was the establishment of technological parameters in order to maximize the elongation of a heat treated special cast iron.

3. Steps for Applying Mathematical Modelling by Active Experiment Using the 2^k Full Factorial Design

The designed experiment is solved according to the following steps [1-9]:

(1) The number of factors (z_i) taken into account is determined and the value of the factors expressed in natural units and coded units is correlated;

(2) The baseline (z_{i0}), the variation range (Δz_i) and the upper (+1) and lower (−1) levels of the analysed factors are determined;

(3) The form of the mathematical model used is established based on the encoded values and taking into account the fact that it's important in optimizing the mathematical model by the design of experiment method, and in this case we can only determine the direction of movement to the optimal range (from a known point) and, therefore, only the linear part of the mathematical model is studied, according to the [1, 2] expression:

$$\tilde{y} = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + b_3 \cdot x_3 + \dots + b_i \cdot x_i \tag{1}$$

(4) The matrix of the 2^k full factorial experiment design is built;

(5) The number of parallel determinations (n_i) to be performed is determined;

(6) The coefficients of the (b_i) mathematical model are calculated using the expression [1, 2]:

$$b_i = \frac{\sum_{i=1}^{N} x_{iu} \cdot \bar{y}_u}{N} \tag{2}$$

where:

x_i = process factors (normalized variable);

$\bar{y}_u$ = arithmetic mean of process performance (arithmetic mean of parallel determinations);

N = number of experimental points (number of rows in the design matrix);

(7) The dispersion of the parallel determinations (row dispersion) at each experimental point (S_u^2) is calculated using the expression [1, 2]:

$$S_u^2 = \frac{\sum_{i=1}^{m} \Delta y_{uk}^2}{m-1} \tag{3}$$

where:

$$\Delta y_{uk}^2 = \left(y_{uk} - \bar{y}_u\right)^2 \tag{4}$$

$$\bar{y}_u = \frac{1}{m} \sum_{k=1}^{m} y_{uk} \tag{5}$$

y_{uk} = state variable in parallel determinations;

m = number of parallel determinations;

(8) The homogeneity of the experimental dispersions is verified using the Cochran criterion [1, 2]:

$$G_C = \frac{\left(S_u\right)_{\max}^2}{\sum_{u=1}^{N} S_u^2} \tag{6}$$

where:

G_C = calculated value of the Cochran criterion;

$\left(S_u\right)_{\max}^2$ = maximum value (from the experiment design matrix) of the experimental dispersions;

The calculated value of the Cochran criterion (G_C) is compared to the critical (tabular) value of the Cochran criterion (G_T), whose expression is:

$$G_T = G_{\alpha;\nu 1;\nu 2} \tag{7}$$

where:

α = statistical coefficient of the confidence level used, α = 0.05;

ν_1, ν_2 = degrees of freedom; $\nu_1 = m - 1$; $\nu_2 = N$;
N= number of experimental points (number of rows in the design matrix).
Two situations arise in this case:
a) $G_C<G_T$, it follows that the experimental dispersions are homogeneous;
b) $G_C>G_T$, it follows that the experimental dispersions are not homogeneous;
If the $G_C<G_T$ rule is complied with, the calculation of the experimental error (reproducibility dispersion), S_0^2 is performed next according to the expression [1, 2]:

$$S_0^2 = \frac{\sum_{u=1}^{N} S_u^2}{m-1} \tag{8}$$

where:
S_u^2 = dispersion of parallel determinations at each experimental point;
m = number of parallel determinations;
(9) The significance of each coefficient "b_i" with a confidence interval "Δb_i" is verified according to the calculation method of the Student criterion.
The confidence interval "Δb_i" is calculated according to the expressions [1, 2]:

$$\Delta b_i = t_{\alpha;\nu_0} \cdot S_{b_i}\,; \tag{9}$$

$$S_{b_i} = \sqrt{S_{b_i}^2}\,; \tag{10}$$

$$S_{b_i}^2 = \frac{S_0^2}{N}\,, \tag{11}$$

where:
S_{b_i} = mean square deviation of "b_i" factors;
$S_{b_i}^2$ = "b_i" factors dispersion.
Two situations arise in this case as well [1-9]:
a) if $|b_i| \geq |\Delta b_i|$, the "b_i" coefficients are part of the relevant mathematical model;
b) if $|b_i| \leq |\Delta b_i|$, the "b_i" coefficients cannot be part of the relevant mathematical model. The calculation is stopped and the coefficient(s) that does (do) not verify the Student criterion is (are) removed;
(9) The correlation between the mathematical model and the experimental data is verified using the Fischer criterion. The possibility of using this mathematical model for optimizing the process is verified on this occasion. This verification is employed to determine whether the calculated approximation of the linear dependence $y = f(x_i)$ is sufficiently precise in relation to the research accuracy. The verification is carried out using the following expressions [1-9]:

$$F_c = \frac{S_{conc}^2}{S_0^2}\,, \tag{12}$$

where:
F_c = calculated value of the Fischer criterion;
S_{conc}^2 = concordance dispersion (error due to the mathematical model);

$$S_{conc}^2 = \frac{m\sum_{i=1}^{N} \Delta\tilde{y}_{ui}^2}{\nu_{conc}}\,, \tag{13}$$

where:
ν_{conc} = number of degrees of freedom used to calculate S_{conc}^2; $\nu_{conc} = N - l$;
l = number of mathematical model coefficients;

and $\Delta\tilde{y}_{ui}^2$ is determined by the expression:

$$\Delta\tilde{y}_{ui}^2 = \left(\bar{y}_u - \tilde{y}_u\right)^2, \qquad (14)$$

where:

$\bar{y}_u$ = arithmetic mean of process performance (arithmetic mean of parallel determinations);

$\tilde{y}_u$ is the calculated state variable (mathematical model value) for each experimental point (for each row in the design matrix);

The calculated value of the Fischer criterion (F_c) will be compared with the critical (tabular) value of the Fischer criterion (F_T) whose expression is:

$$F_T = F\left(\alpha; \nu_{conc}; \nu_0\right), \qquad (15)$$

where:

α = statistical coefficient of the confidence level used, $\alpha = 0.05$;

ν_{conc}; ν_0 = number of degrees of freedom used to calculate the critical (tabular) value of the Fischer criterion (F_T); $\nu_{conc} = N - l$ and $\nu_0 = (m - 1)$, [1-9],

where:

N = number of experimental points (number of rows in the design matrix);

l = number of coefficients in the mathematical model; $l = 4$ (b_0, b_1, b_2, b_3);

n_i = number of parallel determinations.

Two situations arise in this case as well:

a) if $F_c \leq F_T$, this mathematical model matches the experimental data and, therefore, it can be used to determine the direction of movement to an optimal range from a known point;

b) if $F_c \geq F_T$, this mathematical model does not match the experimental data and, therefore, it is not a linear model and it cannot be used in this case as it does not meet the objective of the research.

4. Experimental Procedure

The studied material was a Cu-Ni cast iron with the following composition (% in weight): 3.63 %C; 2.88 %Si; 0.45 %Mn; 0.012 %P; 0.006 %S; 0.050 %Mg; 0.42 %Cu and 0.40 %Ni.

The data presented in this paper is part of a comprehensive study on heat treated SG cast iron hardness. For this purpose, eight specimens, Ø30×5 mm, were used.

The parameters specific to the thermal treatment applied are as follows:

- the austenitizing temperature, t_A[°C];
- the holding time at the austenitizing temperature, τ_A [min];
- the temperature at isothermal level, t_{iz} [°C];
- the holding time at the isothermal level, τ_{iz} [min].

All these experimental specimens, were performed at isothermal maintenance in salt-bath (55% KNO_3+45% $NaNO_3$), and the cooling after the isothermal holding was done in air.

5. Solving the Designed Experiment

The designed experiment is solved according to the steps presented above [1-9].

It was determined as the number of the z_i= 3 factors analysed, i.e. t_A; t_{iz}; τ_{iz}. Table 1 shows the factors analysed alongside base levels and variation ranges.

Table 1. Determination of the factors analysed, of base levels and of variation ranges

Factors	z_1	z_2	z_3
Code	x_1	x_2	x_3
Base level, (z_{i0})	865	350	30
Variation range, (Δz_i)	35	50	20
Upper level, (+1)	900	400	50
Lower level, (-1)	830	300	10

Table 2 shows the 2^3 full factorial experiment design matrix with the results obtained according to the design.

Table 2. Matrix of the 2^3 full factorial experiment design

No.exp.	The order of the experiments	x_i variables				Process performance, y_{uk}			$\bar{y}_u$
		x_0	x_1	x_2	x_3	y_1	y_2	y_3	
1	4	+1	+1	+1	+1	337	325	331	331
2	3	+1	-1	+1	+1	306	311	302	306.3333
3	8	+1	+1	-1	+1	426	415	415	418.6667
4	5	+1	-1	-1	+1	355	344	363	354
5	7	+1	+1	+1	-1	321	331	323	325
6	2	+1	-1	+1	-1	311	306	305	307.3333
7	1	+1	+1	-1	-1	440	451	446	449.3333
8	6	+1	-1	-1	-1	395	398	390	395.3333

The expressions (1÷8) present the calculation of the coefficients of the mathematical model (b_i):

$$b_0 = \frac{331 + 306.3333 + 418.6667 + 354 + 325 + 307.3333 + 449.3333 + 395.3333}{8} = 360.875;$$

$$b_1 = \frac{331 - 306.3333 + 418.6667 - 354 + 325 - 307.3333 + 449.3333 - 395.3333}{8} = 20.125;$$

$$b_2 = \frac{331 + 306.3333 - 418.6667 - 354 + 325 + 307.3333 - 449.3333 - 395.3333}{8} = -43.4583;$$

$$b_3 = \frac{331 + 306.3333 + 418.6667 + 354 - 325 - 307.3333 - 449.3333 - 395.3333}{8} = -8.375;$$

Table 3 shows the calculation of the dispersion of parallel determinations (S_u^2) in each experimental point.

Table 3. Calculation of the dispersion of parallel determinations (S_u^2) in each experimental point

No.exp.	$\bar{y}_u$	Δy_{u1}^2	Δy_{u2}^2	Δy_{u3}^2	ν	S_u^2
1	331	36	36	0	3-1=2	523.333 / 2 = **264.6667**
2	306.3333	0.1111		18.7777		
3	418.6667	**53.7777**	21.7777	13.4444		
4	354	1	13.4444	81		
5	325	16	100	4		
6	307.3333	13.4444	36	5.4444		
7	449.3333	2.7777	1.7777	11.1111		
8	395.3333	0.1111	2.7777	28.4444		
Σ	-	123.2222	243.8889	162.2222	-	-
$\Sigma(\Delta y_u^2)$	-		523.3333		-	-

The homogeneity of the experimental dispersions is verified using the Cochran criterion, according to the expression (6):

$$G_C = \frac{53.7777}{264.6667} = 0.2032 \qquad (16)$$

The critical (tabular) value of the Cochran criterion (G_T) is determined from the criterion-specific tables [1, 2] and the following value is obtained: $G_T = G_{\alpha, \nu, N}$; $G_T = G_{0.05;\,2;\,8} = 0.5157$.

Since $G_C < G_T$ (0.2032 < 0.5157), it follows that the experimental dispersions are homogeneous and the experimental error (the reproducibility dispersion), S_0^2, is calculated next according to the expression (8):

$$S_0^2 = \frac{264.6667}{8} = 33.0833$$

The calculation of the dispersion in determining the coefficients of the mathematical model, calculated using the Student criterion, according to the expressions (9)-(11):

$$S_{bi}{}^2 = \frac{33.0833}{8} = 4.1354\,;\;\; S_{bi} = \sqrt{4.1354} = 2.0336\,;\;\; t_{\alpha\, N} = t_{0.05;\,8} = 2.306\ [2];$$

it follows that: Δb_i = 2.306 2.0336 = 4.6895.

As $|b_i| \geq |\Delta b_i|$: 360.875 ≥ 4.6895 (for b_0); 20.125 ≥ 4.6895 (for b_1); $|-43.4583| \geq |4.6895|$ (for b_2); $|-8.375| \geq |4.6895|$ (for b_3); "b_i" coefficients are part of the mathematical model obtained.

The mathematical model specific to the parameters previously established and according to the expression (1) will take the form: $\tilde{y} = 360.875 + 20.125 \cdot x_1 - 43.4583 \cdot x_2 - 8.375 \cdot x_3$.

The correlation between the mathematical model and the experimental data is verified using the Fischer criterion, according to expressions (12) ÷ (15), as shown in Table 4.

Table 4. The correlation between the mathematical model and the experimental data is verified according to the Fischer criterion

No.exp.	$\bar{y}_u$	$\tilde{y}_u$	$\Delta\tilde{y}_u$	$\Delta\tilde{y}_u^2$	ν_1
1	331	329.1667	1.8333	3.3611	8-4=4
2	306.3333	288.9167	17.4167	303.3403	
3	418.6667	416.0833	2.5833	6.6736	
4	354	375.8333	-21.8333	476.6944	
5	325	345.9167	-20.9167	437.5069	
6	307.3333	305.6667	1.6667	2.7778	
7	449.3333	432.8333	16.5	272.25	
8	395.3333	392.5833	2.75	7.5625	
$\Sigma(\Delta\tilde{y}_u^2)$	-	-	-	1510.167	-

$$S_{conc}^2 = \frac{1510.167}{4} = 377.5417\,;\;\; F_c = \frac{377.5417}{33.0833} = 11.4118$$

The critical (tabular) value of the Fischer criterion (F_T) is determined from the criterion-specific tables [1, 2] and the following value is obtained: $F_T = F(\alpha;\ \nu_1;\ \nu_2) = F(0.05;\ 4;\ 2) = 19.25$.

As $F_c \leq F_T$ (11.4118 ≤ 19.25), this mathematical model matches the experimental data and, therefore, it can be used to determine the direction of movement to an optimal range from a known point.

Following the analysis of the values obtained in the case of the 2^3 full experiment design, presented in Table 4, it is noted that an optimal (maximum) value of the hardness of the analysed material (432.8333 HB) was obtained in the case of the designed experiment no.7, which had the following factors of the analysed process:

x_1 = at the upper level (+1) = 900 °C, corresponding to the austenitizing holding temperature;

x_2 = at the lower level (-1) = 300 °C, corresponding to the isothermal stage holding temperature;

x_3 = at the lower level (-1) = 10 min., corresponding to the isothermal stage holding time.

6. Conclusion

The analysis of all data taken into account leads to the following conclusions:

a) the equation of the determined mathematical model shows that, in the variation ranges chosen for the analysed factors, their influences on the studied process are different, as follows:

- the factor z_1 (b_1 = 20.125) corresponding to the austenitizing holding temperature has the

strongest influence on the process;
- the factor z_3 (b_3 = -8.375) corresponding to the isothermal stage holding time influences the process to a lower extent;
- the factor z_2 (b_2 = -43.4583) corresponding to the isothermal stage holding temperature has the weakest influence on the process;

(b) the experimental dispersions, verified using the Cochran criterion, are homogeneous;
(c) statistically, all "b_i" coefficients of the mathematical model obtained differ from zero and, thus, they are part of the mathematical model obtained, and the verification was performed using the Student criterion;
(d) the optimal (maximum) hardness of the analysed material (432.8333 HB) was obtained for the designed experiment no.7;
(e) the correlation between the experimental data and the calculated mathematical model was verified and the verification was based on the Fischer criterion.

References

1.Taloi D., Bratu C., Florian E., Berceanu E. (1983): *Optimizarea proceselor metalurgice* (*Optimization of metallurgical processes*). Editura Didactiă şi Pedagogică, Bucureşti, p. 76-96
2. Taloi D. (1987): *Optimizarea proceselor tehnologice. Aplicaţii în metalurgie* (*Optimization of technological processes. Applications in metallurgy*).Editura Academiei Române, Bucureşti, p. 162-184
3. Lundstedt T., Seifert E., Abramo L., Thelin B., Nyström Å., Pettersen J., Bergman R. (1998): *Experimental design and optimization*. Chemometrics and Intelligent Laboratory Systems, ISSN: 0169-7439, Vol. 42, No. 1-2, p. 3-40, https://doi.org/10.1016/S0169-7439(98)00065-3
4. Bingham D.R., Li W. (2002): *A class of optimal robust parameter designs*.Journal of Quality Technology, ISSN: 0022-4065, Vol. 34, No. 3, p. 244-259, https://doi.org/10.1080/00224065.2002.11980156
5. van de Ven P.M. (2008): *On the equivalence of three estimators for dispersion effects in unreplicated two-level factorial designs*. Journal of Statistical Planning and Inference, ISSN: 0378-3758, Vol. 138, No. 1, p. 18-29, DOI: 10.1016/j.jspi.2007.05.026
6. Kukreja A., Chopra P., Aggarwal A., Khanna P. (2011): *Application of Full Factorial Design for Optimization of Feed Rate of Stationary Hook Hopper*. International Journal of Modeling and Optimization, ISSN: 2010-3697, Vol. 1, No. 3, p. 205-209, DOI: 10.7763/IJMO.2011.V1.36
7. ***: http://www.southampton.ac.uk/~sml2/screen_assemble/papers/rld.pdf, Accessed 2018-04-11
8. ***: https://www.stat.washington.edu/pds/stat502/LectureNotes/2k.factorial.intro.pdf, Accessed 2018-04-11
9. ***: https://methodology.psu.edu/ra/most/factorial, Accessed: 2018-04-11