20

A Bivariate Plot Useful in Selecting a Robust Design

Albert Prat and Pere Grima

Universitat Politècnica de Catalunya, Barcelona, Spain

Abstract: A sound engineering practice for improving quality and productivity is to design quality into products and processes. The ideas of *G.*Taguchi about parameter design were introduced in the US some ten years ago. Despite some strong controversy about some aspects of it, they play a vital role in the concept of robustness in the design of industrial products and processes. In this paper, we present a new methodology for designing products and processes that are robust to variations in environmental or internal variables. First, a tentative model for the response as a function of the design and noise factors is assumed. This model is then estimated using a single design matrix and the expected value and variance of the response are calculated over the space of the design factors. Finally, the best setting of the parameters values can be located in a newly developed bivariate plot where the distance to the target is plotted against the variance of the response.

Keywords and phrases: Design of experiments, distance-variance plot, quality improvement, robust product, variation reduction

20.1 Introduction

Faced with the old view of quality control by means of final inspection or the classical one of Statistical Process Control, we can designate as the modern method one that guarantees better economy and quality of a product. This would apply to the design of products such that their features are maintained at a desired level despite adverse factors in their production and utilization.

These products are said to be robust. In this paper, we present a new methodology for the selection of values for the parameters used in the design of robust products.

20.2 Background

Techniques for the robust design of products were first introduced in Japan by the engineer Genichi Taguchi (1984, 1986). The method bearing his name is well known and has been used widely. One of the steps involved in using the Taguchi method is the selection of design factors (parameters). This in essence consists of setting up a plan of experimentation where the value of some quality characteristic (response) is measured for a known combination of design factor values and also the so-called noise factors (that influence the response but cannot be easily controlled at a fixed value). From the results obtained during experimentation, one can determine the set of values for the design factors that elicit satisfactory values of the response, independent of the values taken by noise factors.

The method proposed by Taguchi for obtaining the optimal design factor value is through the use of product matrices in the design of the experimental plan and the utilization of signal-to-noise ratios in the analysis of the results [See Hunter (1985), Kacker (1985), Ross (1988), Taguchi and Phadke (1984)].

Despite being well-received, the contributions of Taguchi towards product design have some controversial aspects. Although in some cases the number of experiments that one plans to carry out is justifiably necessary for obtaining precise information, in most other cases, experimenting on each of the conditions of the product matrix would require an effort and dedication of resources that may be unnecessary and therefore could be reduced. [Shoemaker et al. (1991)]. Also the statistics and the associated procedures used are not very intuitive. It has been demonstrated [Box and Fung (1986)] that the statistical techniques used for analyzing the results obtained are not very adequate. For related discussion see also Kacker et al. (1991), Maghsoodloo (1990) and Tort-Martorell (1985).

20.3 Description of the Problem

We consider 3 types of variables that influence the quality characteristics of a product:

i) Those corresponding to parameters related to the design of the product, whose values are maintained constant at the desired levels. This type of variables is known as "constant design factors" or, simply, "design factors".

ii) Those outside the design of the product that show a certain variability around their average value. Typical examples of this type of variables are: the ambient temperature, the degree of humidity, the voltage of the electricity supply, etc. These variables are known as "external noise factors".

iii) The variables corresponding to parameters related to the design of the product (as in type i), whose nominal value can be selected by the designer. In practice, however, they show a certain variability around the value selected. An example of this type of variable is the value of the resistance that is placed in a circuit: the designer chooses its value, for example 10 Ω, but in practice it will not be exactly 10, but between 9.5 and 10.5 Ω. These variables are known as "design factors affected by internal noise" or simply "internal noise factors".

In the design of a product it will not always be necessary to consider the two types of noise factors. In some cases the external noise factors will suffice, and it is possible to consider all the design factors as constant. In others, it will be sufficient to consider solely the variability of some design factors.

The issue that we wish to addess is the determination of the nominal values of the design factors so that: 1) the response shows minimum variability around its average value, neutralizing as far as possible the variability transmitted by the noise factors and, 2) the average value of the response is as near as possible to its objective value. For solving this problem we outline below a 4-stage methodology in which it is assumed that the functional relationship linking response with the factors that influence it is not known, and it is therefore necessary to deduce an approximation experimentally. If this relationship is known, stages i) and ii) should be skipped, since they attempt to find an approximation to the unknown functional relationship.

20.4 Proposed Methodology

The proposed methodology is based on the concepts developed by G. Taguchi (1984, 1986) for the design of robust products, and on later analysis and proposals on this subject, especially the ones by Box and Jones (1990). It consists of the following steps:

i) Assume a tentative model for the response.

ii) Estimate the model parameters by running a factorial or fractional factorial design. If the model is quadratic, a central composite design is used.

iii) Analyze the estimated model, deducing the expressions for the variance and the expected value of the response.

iv) Calculate the expected value and the variance of the response for a large set of combinations of design factor values. Draw a bivariate plot of the optimal distance to the target versus the variance of the response and obtain the optimum combination of values for the design factors.

The use of a spreadsheet program (on a PC) facilitates the calculations and the plotting of the data. Short descriptions of the steps mentioned above are given in the next sections.

20.4.1 Hypothesis about the model for the response

The objective is to establish a hypothesis about the existing relation between the response and the factors (design factors, and noise factors) which must satisfy the following requirements: i) Explain satisfactorily the behaviour of the response, and ii) Facilitate easy estimation of the parameters and the subsequent analysis. Below we discuss 2 typical cases:

1. The noise factors are exclusively of the "external noise" type: in this case, the transmission of variability of the noise factors to the response can be neutralized (as far as possible) using the interaction of the noise factors with the design factors (Figure 20.1). Thus it is sufficient to use models of the type:

$$y = \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \sum_{i=1}^{n-1} \sum_{k=i+1}^{n} \beta_{ik} x_i x_k + \sum_{j=1}^{m} \gamma_j z_j + \sum_{j=1}^{m-1} \sum_{l=j+1}^{m} \gamma_{jl} z_j z_l + \sum_{i=1}^{n} \sum_{j=1}^{m} \delta_{ij} x_i z_j + \varepsilon \tag{20.1}$$

where y: response, x_i: design factors (constants), z_i: noise factors (random variables) and ε: not explained by the model.

2. The noise factors are exclusively of the internal type: In this case, the transmission of variability to the response can be reduced through the mutual interactions of the design factors, but also by using the possible non-linear relationship between the factors affected by variability and the response (Figure 20.2). In this case, it is advantageous to use models of the type:

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \beta_{ij} x_i x_j + \varepsilon \tag{20.2}$$

where $x_1, \ldots, x_k$ represent design factors affected by internal noise. See Box and Jones (1990) and Jones (1990) for a detailed discussion on this area.

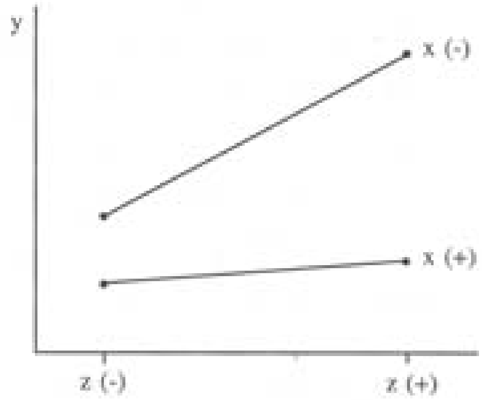

Figure 20.1: The interaction of x with z allows us to choose the values of x (in this case it will be the one coded with '+') which reduces the effect of the variability of z

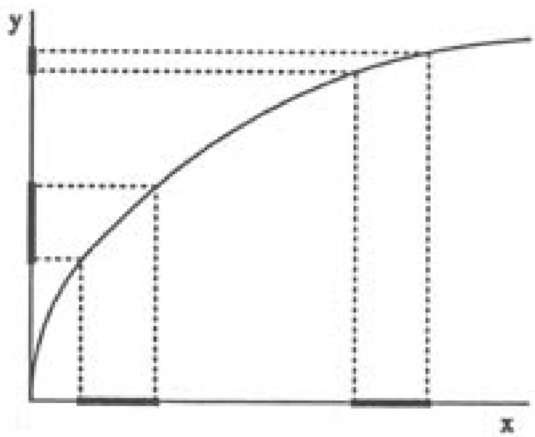

Figure 20.2: The quadratic relation between x and y allows us to reduce the transmission of variability to y by choosing the right values of x (in this case, the greatest possible value)

20.4.2 Estimation of the model parameters

The model parameters can be estimated by running a fractional design. If the model is of first order, a design with 2 levels will be sufficient, but if it is quadratic, it would be necessary to use a central composite design [Box and Draper (1987) and Myers (1976)] or a three level design. In either case, the experimental conditions are formulated throughout using only one design matrix, that we will call a combined matrix, which includes both the control factors and the noise factors. This set-up allows for greater flexibility over the Taguchi product matrix for the estimation of the important effects without confounding. The following example shows this advantage.

Example 1. Let us assume that we have 3 design factors A, B, C and 3 external noise factors O, P and Q. One possible combined matrix is the one that has as generators: $C = AB$ for the inner orthogonal array, and $Q = OP$ for the outer orthogonal array, giving the following defining relation:

$$I = ABC = OPQ = ABCOPQ$$

that corresponds to a fourth of a fraction of a complete factorial design of 2^6 and requires the same number of runs as in the product matrix of Taguchi. From the last defining relation we can see that the main effects cannot be estimated unless we assume that all two factor interactions as non significant. If this assumption cannot be made but we need to estimate the main effects, then the number of experiments required in a Taguchi product matrix is much larger.

However, a combined matrix with the following generators: $P = ABC$ and $Q = BCO$, leads to:

$$I = ABCO = BCOQ = APOQ$$

and consequently we can estimate all the main effects free of two factor interactions. This feature is especially useful when it is not required to estimate all the model parameters. For instance, if we are only interested in minimizing the response variance, all that is required is to estimate the effects of noise factors and the interactions with the control factors. This will be explained in more detail in the next section.

Shoemaker et al. (1989) and Box and Jones (1990) have discussed in detail the advantages of the combined matrix. They also proposed the possibility of using special designs like Addelman or Box-Behnken type designs.

20.4.3 Model analysis: Response variance and expected value

First order models, with only external noise factors

This is a model of the type shown in equation (20.1) and minimising the variability of the response is achieved by minimising the following expression:

$$V(y) = \sum_{j=1}^{m} \left(\gamma_j + \sum_{i=1}^{n} \delta_{ij} x_i \right)^2 V(z_j) + \sum_{j=1}^{m-1} \sum_{l=j+1}^{m} \gamma_{jl}^2 V(z_j z_l) + V(\varepsilon)$$

Considering that ε is independent of any factor, and that all factors included in the model have been previously coded to have values between -1 and $+1$, and making the hypothesis that the factors that can't be controlled (z) have a uniform distribution in that interval, we get the following expression:

$$V(y) = \frac{1}{3} \sum_{j=1}^{m} \left(\gamma_j + \sum_{i=1}^{n} \delta_{ij} x_i \right)^2 + \frac{1}{9} \sum_{j=1}^{m-1} \sum_{l=j+1}^{m} \gamma_{jl}^2 + V(\varepsilon) \qquad (20.3)$$

Example 2. Let us assume that we want to minimise the variability of a quality characteristic (response) that can be expressed by the following model:

$$y = 11 + 2x - 1.5z + 3xz$$

In this case, we have to minimise the following expression:

$$V(y) = (-1.5 + 3x)^2 V(z)$$

The value of x that minimises $V(y)$ is $x = 0.5$.

Example 3. Consider the following model:

$$y = 15 - 5x_1 + 3x_2 - z_1 + z_2 + 2x_1x_2 - 2z_1z_2 + 2.5x_1z_1 + 2.5x_1z_2 + 2x_2z_1 - x_2z_2$$

To minimise the variability of y, we need to minimise the following expression:

$$\begin{aligned} V(y) &= (-1 + 2.5x_1 + 2x_2)^2 V(z_1) + (1 + 2.5x_1 - x_2)^2 V(z_2) + 4V(z_1 z_2) \\ &= (1/3)[(-1 + 2.5x_1 + 2x_2)^2 + (1 + 2.5x_1 - x_2)^2 + 4/9 \end{aligned}$$

Setting the derivatives of $V(y)$ with respect to x_1 and x_2 to zero we obtain: $x_1 = -0.13$ and $x_2 = 0.67$. These will be the optimal values of x_1 and x_2 because any other pair of values would produce a greater variability in the response y.

In the previous examples, the values of the controllable factors which minimise the variability of the response were found. But this was done without taking into account the mean value of the response. However, the response will have an optimal value (that we shall call τ). We will also be interested in minimising the distance between the mean value of the response and its optimal value. Therefore, we have to minimise the following expression:

$$\tau - E(y) = \tau - \beta_0 + \sum_{i=1}^{n} \beta_i x_i + \sum_{i=1}^{n-1} \sum_{k=i+1}^{n} \beta_{ik} x_i x_k$$

where: $E(\varepsilon) = 0$ and $E(z_i) = 0$.

Example 4. Given the simple model in Example 2:

$$y = 11 + 2x - 1.5z + 3xz$$

we want to minimise the distance between the mean value of the response and its optimal value (assuming a value of $\tau = 10$). For this we should find the value of x that minimises (for absolute values) the following expression:

$$D(\tau) = 10 - 11 - 2x$$

It can be easily seen that the optimal value of x is -0.5 because for this value we obtain $D(\tau, \underline{x}) = 0$.

In general, the distance to the optimum will be:

$$D(\tau, \underline{x}) = \tau - E(y) \tag{20.4}$$

Second order models, with only internal noise factors

In this case the model corresponds to the type shown in equation (20.2) and hence we have:

$$V(y) = V\left(\beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \beta_{ij} x_i x_j\right) + V(\varepsilon)$$

When there is only one design factor, a linear approximation of the variation transmitted to the response takes the form:

$$V(y) = V(x)(dy/dx)^2 + V(\varepsilon)$$

In general, with k design factors we obtain:

$$V(y) = \sum_{i=1}^{k} V(x_i)(\partial f/\partial x_i)^2 + V(\varepsilon) \tag{20.5}$$

where the value of $V(\varepsilon)$ may be ignored as it only adds a constant to $V(y)$, and does not affect the conclusions obtained.

Since $E(\varepsilon) = 0$ and the x_i (design factors) can be considered random variables independent of ε, with a distribution: $x_i \sim N(\mu_i, \sigma_i)$, the mathematical expectation for y at any defined point is:

$$E(y) = \beta_0 + \sum_{i=1}^{k} \beta_i E(x_i) + \sum_{i=1}^{k} \beta_{ii} E(x_i^2) + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \beta_{ij} E(x_i) E(x_j) ,$$

and:

$$E(x^2) = V(x) + E^2(x) .$$

Then:

$$E(y) = \beta_0 + \sum_{i=1}^{k} \beta_i E(x_i) + \sum_{i=1}^{k} \beta_{ii} Var(x_i) + \sum_{i=1}^{k} \beta_{ii} E(x_i)^2 + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \beta_{ij} E(x_i) E(x_j)$$

20.4.4 Choosing the optimum values for design factors: the distance-variance plot

As shown previously, the value that minimises the response variability is different from the one that minimises its distance to the optimum.

There is no general rule that can facilitate the selection of an optimal value for $\underline{x}$, because in some cases we would like to minimise the response variability even if there is a large distance to the target and in other cases do the reverse.

Looking at the model used in examples 2 and 4, a way of selecting the value of x consists of first calculating the values of $E(y)$ and $V(y)$ for every possible value of the control factor. Then, on examination of the calculated values, it is possible to select the value of x that best fits the requirements of the designer.

A simple method that can be generalised for more complicated models for the selection of the optimal value of x, is given below. It is possible to calculate and display graphically all the relevant data by using any spreadsheet programme running on a PC. The steps involved are:

1. In column A of the spreadsheet, enter the values of x between the intervals -1 and $+1$ with increments of 0.1 (this increment could be reduced).
2. In column B, calculate the variance of the response for each value of x, using formula (20.3).
3. In column C, calculate the distance to the target using formula (20.5).

An illustration of the method is shown in Table 20.1. The model parameters and the optimal value for the response (τ) are entered in column G, so that if any of the data is changed, $V(y)$ and the distance to the target are automatically re-calculated.

From this spreadsheet tabulation, a bivariate plot of the distance to the target and the variance can be generated as shown in Figure 20.3. Note that every data point plotted, corresponds to a value of x. From Figure 20.3, one can observe that there is a value of x for which the variance equals zero and the distance to the target is equal to one. There is also an x for which the distance equals zero and the variance equals 0.75. The plot in Figure 20.3 has all the information required in order to select the data point that corresponds to the values of the design factors that best fit the objectives of the designer.

The identification of which value of x corresponds to each one of the plotted data points is simple. In this case, it can be done by counting in increments of 0.1 from the data points that correspond to $x = -1$ and $x = +1$. In other cases, they can be estimated either from their coordinates or by using software packages such as STATGRAPHICS that allow for the identification of any characteristic of a plotted point by just placing the cursor on it.

Additionally, as shown above, this kind of plot can be used for any number of design factors and noise factors. Taking the model used in Example 3, there are two design factors that are both between the interval $-1, +1$ with an increment of 0.1. It is possible to have $21 \times 21 = 441$ combinations of these for which we can calculate the response variance and the distance to the optimum using the spreadsheet programme (see Table 20.2).

The distance-variance plot obtained from this computation is shown in Figure 20.4. From the figure we get $\tau = 24$. The plotted data point that corresponds to a minimum variance has been highlighted with a box around it and the values of x for this are $x_1 = -0.1$ and $x_2 = 0.6$.

20.5 Conclusions

The most important aspects of this method may be summarized in five points:

i. When the functional relationship between the response and the factors is not known it can be found through the experimentation using a single design matrix that includes both, noise and design factors. This provides sensible estimates of the effects and a reduction in the number of experiments [In agreement with the proposals of Shoemaker et al. (1991)].

ii. The proposed method does not include abstract concepts, and is easily understood, used and interpreted even by people with little knowledge of mathematics or statistics.

iii. The calculations are simple and a spreadsheet program for personal computers (e.g. LOTUS) is the only tool used in the application of this method.

iv. The information is presented as a graph (it is always a bivariate diagram) that clearly summarizes the information available in order to make the most suitable decision in every different case.

v. The method is valid for the design of robust products subject to either internal or external noise.

Table 20.1: Spreadsheet calculation to obtain the values for variance and the distance to the target for examples 2 and 4

	A	B	C	D	E	F	G	H
1	X	Var(Y)	Dist.					
2	−1	6.75	2			cte	11	
3	−0.9	5.88	1.8			x	2	
4	−0.8	5.07	1.6			z	−1.5	
5	−0.7	4.32	1.4			xz	3	
6	−0.6	3.63	1.2					
7	−0.5	3	1			tau	10	
8	−0.4	2.43	0.8					
9	−0.3	1.92	0.6					
10	−0.3	1.47	0.4					
11	−0.1	1.08	0.2					
12	0	0.75	0					
13	0.1	0.48	−0.2					
14	0.2	0.27	−0.4					
15	0.3	0.12	−0.6					
16	0.4	0.03	−0.8					
17	0.5	0	−1					
18	0.6	0.03	−1.2					
19	0.7	0.12	−1.4					
20	0.8	0.27	−1.6					
21	0.9	0.48	−1.8					
22	1	0.75	−2					

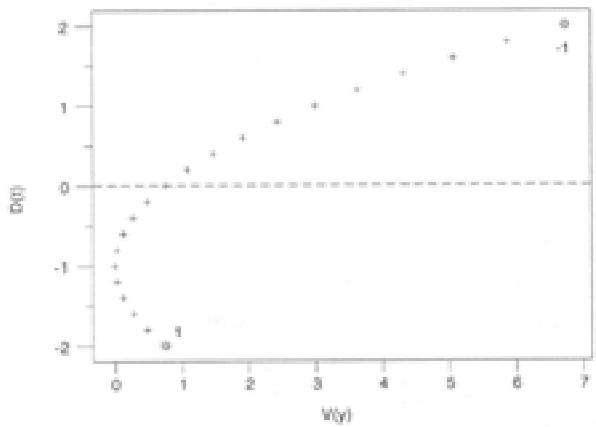

Figure 20.3: Distance-Variance plot for examples 2 and 4

Table 20.2: The first row of the table of values for the model in example 3

	A	B	C	D	E	F	G
1	$x1$	$x2$	$V(y)$	Dist.			
2	−1	−1	10.6	5		cte	15
3	−1	−0.9	9.9	4.9		$x1$	−5
4	−1	−0.8	9.3	4.8		$x2$	3
5	−1	−0.7	8.7	4.7		$z1$	−1
6	−1	−0.6	8.1	4.6		$z2$	1
7	−1	−0.5	7.5	4.5		$x1x2$	2
8	−1	−0.4	7.0	4.4		$z1z2$	−2
9	−1	−0.3	6.5	4.3		$x1z1$	2.5
10	−1	−0.2	6.1	4.2		$x1z2$	2.5
11	−1	−0.1	5.7	4.1		$x2z1$	2
12	−1	0	5.3	4		$x2z2$	−1
13	−1	0.1	4.9	3.9			
14	−1	0.3	4.6	3.8		tau	24
15	−1	0.4	4.3	3.7			

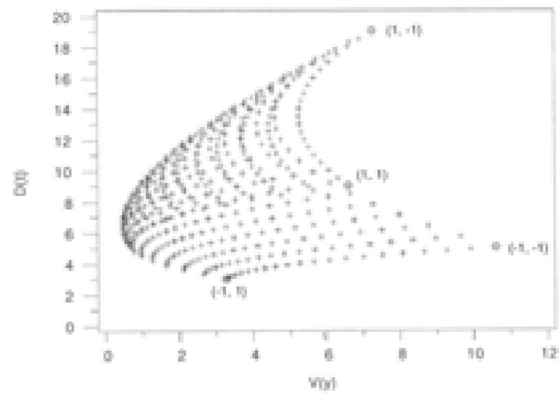

Figure 20.4: Distance-variance plot for example 3

References

1. Box, G. E. P., Hunter W. G. and Hunter J. S. (1978). *Statistics for experimenters*, New York: John Wiley.

2. Box, G. E. P. and Fung, C. A. (1986). Studies in quality improvement: minimizing transmitted variation by parametrer design, *Center for Quality and Productivity Improvement*, Report **8**, University of Wisconsin-Madison.

3. Box, G. E. P. and Draper N. R. (1987). *Empirical Model Building and Response Surfaces*, New York: John Wiley.

4. Box G. E. P. and Jones S. (1990). Designing products that are robust to the environment, *Center for Quality and Productivity Improvement*, Report **28**, University of Wisconsin-Madison.

5. Hunter, J. S. (1985). Statistical design applied to product design, *Journal of Quality Technology*, **17**, 210–221.

6. Jones, S. P. (1990). Designs for minimizing the effect of envirinmental variables, *Unpublished Ph.D. dissertation*, University of Wisconsin, Department of Statistics.

7. Kackar, R. N. (1985). Off-Line quality control, parametrer design, and the Taguchi method, *Journal of Quality Technology*, **17**, 176–188.

8. Kackar, R. N., Lagergren, E. S. and Filliben. J. J. (1991). Taguchi's fixed-element arrays are fractional factorials, *Journal of Quality Technology*, **23**, 107–115.

9. Maghsoodloo, S. (1990). The exact relation of Taguchi's signal-to-noise ratio to his quality loss function, *Journal of Quality Technology*, **22**, 57–67.

10. Myers, R. H. (1976). *Response surface methodology*, Boston: Allyn and Bacon, Inc.

11. Ross, P. J. (1988). *Taguchi techniques for quality engineering*, New York: McGraw-Hill.

12. Shoemaker, A. C., Tsui K. L. and Wu, C. F. J. (1991). Economical experimentation methods for robust design, *Technometrics*, **33**, 415–427.

13. Taguchi G. and Phadke M. S. (1984). Quality engineering through design optimization, *Presented at IEEE GLOBECOM, Atlanta.*

14. Taguchi, G. (1986). *Introduction to quality engineering: designing quality into products and processes*, Tokyo: Asian Productivity Organization.

15. Taguchi, G., Elsayed, E. A. and Hsiang, T. (1989). *Quality engineering in production systems*, New York: McGraw-Hill.

16. Tort-Martorell, J. (1985). Diseños factoriales fraccionales, aplicación al control de calidad mediante el diseño de productos y procesos, *Unpublished Ph.D. dissertation*, Universitat Politècnica de Catalunya, Spain.