OPEN ACCESS

International Journal Of Modern Engineering Research (IJMER)

Robust Design of Experiments: A Methodological Approach

A. Bizimana[1], J. D. Domínguez[2], H. VaqueraH.[3]

[1](Colegio De Postgraduados –Campus Montecillo, México)
[2](Centro De Investigación En Matemáticas –Unidad Aguascalientes, México)
[3](Colegio De Postgraduados –Campus Montecillo, México)

ABSTRACT: *We provide a methodology for conducting robust design and analysing the data obtained from the experiment. The methods we expose are the Taguchi approach, commonly known as double orthogonal array design (DOAD), and the combined array design (CAD). The Taguchi approach consists of a double orthogonal array design, one for the design factors and another for the noise factors. The combined array design puts both types of factors in one design. This design permits the analysis of interactions between the design and noise factors, and reduces the number of runs required to conduct an experiment. The analysis of data obtained from this design consists of fitting a regression model in terms of design factors and noise factors. From the adjusted model, two response surfaces are obtained, one for the mean of the quality characteristic and another for its variance. Optimal conditions of the experiment are determined by solving the optimization problems which are based on mean square error (MSE) criterion and desirability function (DF). Based on quadratic loss function, we assess the impact of a robust design in reducing the production cost.*

Keywords: *Robust design, double orthogonal array design, combined array design, mean square error, desirability function, quadratic function.*

I. INTRODUCTION

Since engineers and scientists have become increasingly aware of the benefits of using designed experiments, there have been many new areas of application. One of the most important is the robust design. Robust design methodology is a systematic effort to achieve insensitivity to noise factors.

The assumption is that there are two types of factors that affect the quality characteristic commonly known as response variable. These are the control factors and uncontrollable or difficult-to-control factors. They are respectively referred to as design factors and noise factors (Taguchi [1]).

Noise factors can be further divided into two categories: external noise factors and internal noise factors. External noise factors are those sources of variability that come from outside of the system. Examples of external noise factors are environmental factors that a system is subject to, such as ambient temperature, ambient pressure and humidity. Internal noise factors are essentially from the variations of control factors. Internal noise could include deviations from the target values of control factors caused by the manufacturing process, assembly and deterioration.

Robust design is mainly composed of three stages: robust system design, robust parameter design and robust tolerance design (Harrington [2]). Robust system design consists of using physics, mathematics, experience and knowledge gained in a specific field to develop and select the most appropriate conditions of the design. Once the configuration of a system is finalised, the settings of the nominal levels and the corresponding tolerances need to be determined. Robust parameter design aims at finding the optimal settings of control factors so that the system is insensitive or less sensitive to noise factors. Robust tolerance design is a balancing process. It aims at finding the optimal settings of tolerances of the control factors so that the total cost of the system is minimal (Harrington [2]).

The statistical methodology underlying robust design that has by now become the most widely accepted, is the dual response surface methodology which estimates two surfaces, one for the mean and another for the variance of the quality characteristic (Giovagnoli and Romano[3]). As an illustration, we adopt the data analysis based upon the double orthogonal array design and combined array design. Details on these experimental strategies are in (Bizimana [4]).

In literature, much has been done on robust design. This paper presents a methodology for conducting robust design and analysing the data obtained from the experiment using two experimental schemes. From there we build the dual response strategy for both schemes. The models for the mean and variance require an optimization process. This process is based upon the mean square error and desirability function. The proposed methodology is described in Table 1. Finally, an example from the literature is used to study the efficiency of the dual response in both designs. The comparison of the procedures for optimization reveals that the combined array design produces better results. Optimal conditions are applied to the quadratic loss function, generating an economic impact.

Table 1: Proposed methodology.

Design	Model	Dual Response	Optimization 1	Optimization 2	Application
DOAD	Fixed effects	Mean: $\hat{y}_{Mean}$ Variance: $\hat{y}_{Var}$	MSE	Desirability function	Loss function
CAD	Mixed effects	Mean: $E(\hat{y})$ Variance: $Var(\hat{y})$	MSE	Desirability function	Loss function

II. EXPERIMENTAL SCHEMES

2.1. Double orthogonal array design

The double orthogonal array design was initiated by Genichi Taguchi (Taguchi and Wu [5]). It consists of a cross-product of two experimental designs. The first design, known as the inner design, is a combination of the levels of the design factors. The second design, referred to as the outer array design, is a combination of the levels of the noise factors. Each combination of the levels of the design factors forms an experimental run. For each experimental run, the same array of the noise factors is run.

Suppose that the quality characteristic y of a product or a process depends on p design factors $x_1,\ldots,x_p$ and q noise factors $z_1,\ldots,z_q$. The responses y_{ij} are the combinations of the levels of the design factors $(i=1,2,\cdots,n)$ and the levels of the noise factors $(j=1,2,\cdots,r)$. The total number of runs required to conduct an experiment in this case is $n\times r$. The experimental structure of double orthogonal array design is represented by Figure 1.

2.2. Combined Array Design

The combined array design is a single experimental design in control and noise factors. Both control and noise factors are then modeled. The results of the experiment can be described by a model with only a small number of main effects and low-order interactions. Significant design-by-noise interactions are interpreted as evidence of dispersion effects and are used to choose settings of design factors that minimize the process variation. Data obtained from combined array design are analysed by fitting a model for the mean and the variance.

Figure 1: Experimental structure of double orthogonal array design.

			z_1	z_{11}	z_{21}	$\cdots$	z_{r1}		
			z_2	z_{12}	z_{22}	$\cdots$	z_{r2}		
			$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		
			z_q	z_{1q}	z_{2q}	$\cdots$	z_{rq}		
x_1	x_2	$\cdots$	x_p	Observations				$\bar{y}$	s^2
x_{11}	x_{12}	$\cdots$	x_{1p}	y_{11}	y_{12}	$\cdots$	y_{1r}	$\bar{y}_1$	s_1^2
x_{21}	x_{22}	$\cdots$	x_{2p}	y_{21}	y_{22}	$\cdots$	y_{2r}	$\bar{y}_2$	s_2^2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_{n1}	x_{n2}	$\cdots$	x_{np}	y_{n1}	y_{n2}	$\cdots$	y_{nr}	$\bar{y}_n$	s_n^2

Suppose that the quality characteristic y of a product or process depends on p design factors $x_1,\ldots,x_p$ and q noise factors $z_1,\ldots,z_q$. The experimental structure of the combined array design is presented by Figure 2.

Figure 2: Experimental structure of combined array design.

x_1	x_2	$\dots$	x_p	z_1	z_2	$\dots$	z_q	y
x_{11}	x_{12}	$\dots$	x_{1p}	z_{11}	z_{12}	$\dots$	z_{1q}	y_{11}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_{n1}	x_{n2}	$\dots$	x_{np}	z_{r1}	z_{r2}	$\dots$	z_{rq}	y_{nr}

III. DUAL RESPONSE OPTIMIZATION

3.1. Dual response in double orthogonal array design

The data analysis consists of fitting a second order regression model of the form

$$y = \beta_0 + \boldsymbol{x}^T \boldsymbol{\beta} + \boldsymbol{x}^T \boldsymbol{B}\,\boldsymbol{x} + \varepsilon. \qquad (1)$$

In this model, $\boldsymbol{x}$ is the vector of control factors, β_0 the intercept, $\boldsymbol{\beta}$ is a vector of coefficients of first-order control factors, $\boldsymbol{B}$ is a matrix of coefficients of second-order terms of control factors and their interactions, $\boldsymbol{\varepsilon}$ is a vector of residual errors of the regression model. The residual errors are assumed to be $N(0,\sigma^2)$. The response y is the mean or the variance.

After the estimate of the regression model $\hat{y} = \hat{\beta}_0 + \boldsymbol{x}^T \hat{\boldsymbol{\beta}} + \boldsymbol{x}^T \hat{\boldsymbol{B}} \boldsymbol{x}$ is obtained, the mean response surface, and variance response surface are $\hat{y}_{Mean}(\boldsymbol{x})$ and $\hat{y}_{Var}(\boldsymbol{x})$, respectively. It is clear that both response surfaces are in terms of control factors.

3.2. Dual response in combined array design

Let the system be described by a variable $y(\boldsymbol{x},\boldsymbol{z})$ that depends on a set of controllable factors (the vector $\boldsymbol{x}$) and a set of random noise factors (the vector $\boldsymbol{z}$).

To explore the dependence of y on $\boldsymbol{x}$ and $\boldsymbol{z}$, the following model is assumed for the response, to accommodate control-by-noise interactions:

$$y(\boldsymbol{x},\boldsymbol{z}) = \beta_0 + \boldsymbol{x}^T \boldsymbol{\beta} + \boldsymbol{x}^T \boldsymbol{B}\,\boldsymbol{x} + \boldsymbol{\gamma}^T \boldsymbol{z} + \boldsymbol{x}^T \boldsymbol{\Delta}\,\boldsymbol{z} + \varepsilon \qquad (2)$$

In this model, $\boldsymbol{z}$ is the random noise vector, the ε's are independent and identically distributed $N(0,\sigma^2)$ random errors. It is assumed that $\boldsymbol{\varepsilon}$ and $\boldsymbol{z}$ are independent. The constant β_0, the vectors $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and the matrices $\boldsymbol{B}$ and $\boldsymbol{\Delta}$ consist of unknown parameters, and σ^2 is also usually unknown. It is also assumed that $E(\boldsymbol{z}) = \boldsymbol{0}$ and that $Cov(\boldsymbol{z}) = \boldsymbol{\Omega}$ is known.

After the model (2) is fitted to the data from the designed experiment, the corresponding adjusted response model is given by the expression

$$\widehat{y(\boldsymbol{x},\boldsymbol{z})} = \hat{\beta}_0 + \boldsymbol{x}^T \hat{\boldsymbol{\beta}} + \boldsymbol{x}^T \hat{\boldsymbol{B}}\,\boldsymbol{x} + \hat{\boldsymbol{\gamma}}^T \boldsymbol{z} + \boldsymbol{x}^T \hat{\boldsymbol{\Delta}}\,\boldsymbol{z} \qquad (3)$$

The two response surfaces are obtained analytically from (3). Both response surfaces are in terms of control factors. The mean response surface, $E_z\left(\widehat{y(\boldsymbol{x},\boldsymbol{z})}\right)$, and the variance response surface, $Var_z\left(\widehat{y(\boldsymbol{x},\boldsymbol{z})}\right)$, are respectively expressed as follows:

$$E_z\left(\widehat{y(\boldsymbol{x},\boldsymbol{z})}\right) = \hat{\beta}_0 + \boldsymbol{x}^T \hat{\boldsymbol{\beta}} + \boldsymbol{x}^T \hat{\boldsymbol{B}}\,\boldsymbol{x} \qquad (4)$$

$$Var_z\left(\widehat{y(\boldsymbol{x},\boldsymbol{z})}\right) = \left(\hat{\boldsymbol{\gamma}}^T + \boldsymbol{x}^T \hat{\boldsymbol{\Delta}}\right)\boldsymbol{\Omega}\left(\hat{\boldsymbol{\gamma}} + \hat{\boldsymbol{\Delta}}^T \boldsymbol{x}\right). \qquad (5)$$

In fact, calculating the expected value and the variance of $\widehat{y(\boldsymbol{x},\boldsymbol{z})}$ with respect to the random vector $\boldsymbol{z}$ leads to

$$\begin{aligned}
E_z\left(\widehat{y(\boldsymbol{x},\boldsymbol{z})}\right) &= E_z\left[\hat{\beta}_0 + \boldsymbol{x}^T \hat{\boldsymbol{\beta}} + \boldsymbol{x}^T \hat{\boldsymbol{B}}\boldsymbol{x} + \hat{\boldsymbol{\gamma}}^T \boldsymbol{z} + \boldsymbol{x}^T \hat{\boldsymbol{\Delta}}\boldsymbol{z}\right] \\
&= E_z\left[\hat{\beta}_0 + \boldsymbol{x}^T \hat{\boldsymbol{\beta}} + \boldsymbol{x}^T \hat{\boldsymbol{B}}\boldsymbol{x}\right] + E_z\left[\hat{\boldsymbol{\gamma}}^T \boldsymbol{z} + \boldsymbol{x}^T \hat{\boldsymbol{\Delta}}\boldsymbol{z}\right] \\
&= E_z\left[\hat{\beta}_0 + \boldsymbol{x}^T \hat{\boldsymbol{\beta}} + \boldsymbol{x}^T \hat{\boldsymbol{B}}\boldsymbol{x}\right] \\
&= \hat{\beta}_0 + \boldsymbol{x}^T \hat{\boldsymbol{\beta}} + \boldsymbol{x}^T \hat{\boldsymbol{B}}\boldsymbol{x},
\end{aligned}$$

and

$$\begin{aligned} Var_z\left(\widehat{y(\boldsymbol{x},\boldsymbol{z})}\right) &= Var_z\left[\widehat{\beta}_0+\boldsymbol{x}^T\widehat{\boldsymbol{\beta}}+\boldsymbol{x}^T\widehat{\boldsymbol{B}}\boldsymbol{x}+\widehat{\boldsymbol{\gamma}}^T\boldsymbol{z}+\boldsymbol{x}^T\widehat{\boldsymbol{\Delta}}\boldsymbol{z}\right] \\ &= Var_z\left[\widehat{\beta}_0+\boldsymbol{x}^T\widehat{\boldsymbol{\beta}}+\boldsymbol{x}^T\widehat{\boldsymbol{B}}\boldsymbol{x}\right]+Var_z\left[\widehat{\boldsymbol{\gamma}}^T\boldsymbol{z}+\boldsymbol{x}^T\widehat{\boldsymbol{\Delta}}\boldsymbol{z}\right] \\ &= Var_z\left[\widehat{\boldsymbol{\gamma}}^T\boldsymbol{z}+\boldsymbol{x}^T\widehat{\boldsymbol{\Delta}}\boldsymbol{z}\right] \\ &= Var_z\left[\left(\widehat{\boldsymbol{\gamma}}^T+\boldsymbol{x}^T\widehat{\boldsymbol{\Delta}}\right)\boldsymbol{z}\right] \\ &= \left(\widehat{\boldsymbol{\gamma}}^T+\boldsymbol{x}^T\widehat{\boldsymbol{\Delta}}\right)Var(\boldsymbol{z})\left(\widehat{\boldsymbol{\gamma}}^T+\boldsymbol{x}^T\widehat{\boldsymbol{\Delta}}\right)^T \\ &= \left(\widehat{\boldsymbol{\gamma}}^T+\boldsymbol{x}^T\widehat{\boldsymbol{\Delta}}\right)\boldsymbol{\Omega}\left(\widehat{\boldsymbol{\gamma}}+\widehat{\boldsymbol{\Delta}}^T\boldsymbol{x}\right). \end{aligned}$$

3.3. Optimal dual response

In the literature, various methods of optimization have been developed in order to obtain the optimal solution for the mean of the quality characteristic while minimizing the variance of the process. Mares and Domínguez [6] have summarized and compared those methods. Myers and Carter [7], and Myers and Vining [8] have introduced the method commonly used in the dual response surface approach. They first fit second order models to both primary and secondary response surfaces. In this case, they are respectively, the mean and variance. Then, they optimize the primary response subject to an appropriate constraint on the value of the secondary response, or vice versa.

The optimal solution for the mean response is obtained by solving the problem:

$$\begin{cases} \text{Optimize} & \widehat{y}_{Mean}(\boldsymbol{x}) \\ \text{Subject to} & \widehat{y}_{Var}(\boldsymbol{x})=\sigma_0^2 \qquad \text{(Case of DOAD)} \\ & \boldsymbol{x}\in R \end{cases} \tag{6}$$

and

$$\begin{cases} \text{Optimize} & E_z\left(\widehat{y(\boldsymbol{x},\boldsymbol{z})}\right) \\ \text{Subject to} & Var_z\left(\widehat{y(\boldsymbol{x},\boldsymbol{z})}\right)=\sigma_0^2 \qquad \text{(Case of CAD)} \\ & \boldsymbol{x}\in R. \end{cases} \tag{7}$$

Lin and Tu [9] argue that the methods of optimization in Equations (6) and (7) may be misleading, because the variance, which is to be minimized in the process, is forced to a fixed value. They propose a new procedure based on the mean square error criterion.

The optimal solution for the variance model is the solution of the following problem:

$$\begin{cases} \text{Minimize} & \widehat{y}_{Var}(\boldsymbol{x}) \\ \text{Subject to} & \widehat{y}_{Mean}(\boldsymbol{x})=T \qquad \text{(Case of DOAD)} \\ & \boldsymbol{x}\in R \end{cases} \tag{8}$$

and

$$\begin{cases} \text{Minimize} & Var_z\left(\widehat{y(\boldsymbol{x},\boldsymbol{z})}\right) \\ \text{Subject to} & E_z\left(\widehat{y(\boldsymbol{x},\boldsymbol{z})}\right)=T \qquad \text{(Case of CAD)} \\ & \boldsymbol{x}\in R. \end{cases} \tag{9}$$

R is the experimental region.

3.4. Economic impact of designed experiments

In this section, we assess the economic impact of designed experiment using the quadratic loss function. The quadratic loss function (QLF) is a metric used to provide a better estimate of the monetary loss incurred by manufacturers and consumers when the product performance deviates from its target value (Meng *et al.*, [10]). Then, the QLF is used to evaluate the economic impact of conducting an experiment on the process.

The QLF is given by the expression

$$L(y) = k(y-T)^2 \tag{10}$$

where y is the quality characteristic of a product or process, T is the target and k is the quality loss coefficient.

The expected quality loss (EQL) is

$$Q = E\left[L(y)\right] = k\,E(y-T)^2 = k\left[\left(E(y)-T\right)^2 + Var(y)\right]. \tag{11}$$

By taking $E(y) = \mu$ and $Var(y) = \sigma^2$, the expected quality loss becomes

$$Q = k\left[\left(\mu - T\right)^2 + \sigma^2\right]. \tag{12}$$

Then the estimate of expected quality loss is:

$$\widehat{Q} = k\left[\left(\widehat{\mu} - T\right)^2 + \widehat{\sigma^2}\right], \tag{13}$$

where $\widehat{\mu} = \overline{y}$ and $\widehat{\sigma^2} = s^2$.

The quality loss coefficient k is determined by first finding the functional limits or customer tolerance for y. The function limits are determined by $T \pm \Delta_0$. These are the points at which the product would fail or produce unacceptable performance in approximately half of the customer applications. Let A_0 be the value of the quality loss function at $T \pm \Delta_0$, that is $L(y) = A_0$ at $y = T \pm \Delta_0$. Substituting the functional limits $T \pm \Delta_0$ and the value of the quality loss into Equation (10), the quality loss coefficient is found to be

$$k = \frac{A_0}{(\Delta_0)^2}. \tag{14}$$

3.4.1. Types of quadratic loss function

While conducting an experiment, the designer is interested in reaching the target value or minimizing or maximizing the value of the quality characteristic. These three cases of quality characteristics are referred to as the nominal the best type (NTB), the smaller the better (STB) and the larger the better (LTB), respectively.

Let $y^T = (y_1, y_2, \ldots, y_n)$ where y is the quality characteristic of a product or process. Table 2 shows the types of QLF and the average quadratic loss functions corresponding to each kind of quality characteristic of interest. More details on this section can be found in Fowlkes and Creveling [11].

Table 2: Types of QLF.

Type	QuadraticLoss	EQL	Estimate of EQL
NTB	$L(y) = k(y-T)^2$	$Q = k\left(\left(\mu - T\right)^2 + \sigma^2\right)$	$\widehat{Q} = k\left(\left(\overline{y} - T\right)^2 + s^2\right)$
STB	$L(y) = k\,y^2$	$Q = k\left(\mu^2 + \sigma^2\right)$	$\widehat{Q} = k\left(\overline{y}^2 + s^2\right)$
LTB	$L(y) = \frac{k'}{y^2}$	$Q = \frac{k'}{\mu^2}\left[1 + \frac{3\sigma^2}{\mu^2}\right]$	$\widehat{Q} = \frac{k'}{\overline{y}^2}\left[1 + \frac{3s^2}{\overline{y}^2}\right]$

In this table, $k = \frac{A_0}{(\Delta_0)^2}$ and $k' = A_0\left(\Delta_0\right)^2$.

IV. OPTIMIZATION PROCEDURE

4.1. Mean square error

The mean square error is an effective criterion to combine the mean and the standard deviation responses in dual response optimization (Köksoy [12]).

The MSE function for the target is best case is given by

$$\widehat{MSE} = \begin{cases} \left[\hat{y}_{Mean}(\boldsymbol{x}) - T\right]^2 + \hat{y}_{Var}(\boldsymbol{x}) & \text{(Case of DOAD)} \\ \left[E_z\left(\widehat{y(\boldsymbol{x},\boldsymbol{z})}\right) - T\right]^2 + Var_z\left(\widehat{y(\boldsymbol{x},\boldsymbol{z})}\right) & \text{(Case of CAD)} \end{cases} \tag{15}$$

where T is the target value.

The MSE function for the smaller the better case is determined as follows:

$$\widehat{MSE} = \begin{cases} \left[\hat{y}_{Mean}(\boldsymbol{x})\right]^2 + \hat{y}_{Var}(\boldsymbol{x}) & \text{(Case of DOAD)} \\ \left[E_z\left(\widehat{y(\boldsymbol{x},\boldsymbol{z})}\right)\right]^2 + Var_z\left(\widehat{y(\boldsymbol{x},\boldsymbol{z})}\right) & \text{(Case of CAD)} \end{cases} \tag{16}$$

The MSE function for the larger the better case is given by the expression

$$\widehat{MSE} = \begin{cases} \left[\hat{y}_{Mean}(\boldsymbol{x}) - H\right]^2 + \hat{y}_{Var}(\boldsymbol{x}) & \text{(Case of DOAD)} \\ \left[E_z\left(\widehat{y(\boldsymbol{x},\boldsymbol{z})}\right) - H\right]^2 + Var_z\left(\widehat{y(\boldsymbol{x},\boldsymbol{z})}\right) & \text{(Case of CAD)} \end{cases} \tag{17}$$

where H is the highest plausible value of $\hat{y}_{Mean}(\boldsymbol{x})$ or $E_z\left(\widehat{y(\boldsymbol{x},\boldsymbol{z})}\right)$.

The optimization problem to solve is then

$$\begin{cases} \text{Minimize} \quad \widehat{MSE} \\ \text{Subject to} \quad \boldsymbol{x} \in R. \end{cases} \tag{18}$$

The advantage of the MSE approach is that it does not require any constraints on the secondary response, and it can handle more realistic models and much more complicated models than polynomial.

4.2. Desirability function

The desirability function to simultaneously optimize multiple equations was originally proposed by Harrington [2]. The common approach is to transform each response $\widehat{y_i}$ into an individual function d_i that varies over the range $[0,1]$.

4.2.1. Individual desirability function

Depending on whether a particular response $\widehat{y_i}$ is to be maximized, minimized, or assigned to a target value, the corresponding desirability function, d_i, is defined as follows (Derringer and Suich [13]):

- Individual desirability function for the nominal the best (NTB) case:

$$d_i = \begin{cases} 0 & \text{if } \widehat{y_i(\boldsymbol{x})} \le L_i \text{ or } \widehat{y_i(\boldsymbol{x})} \ge U_i \\ \left(\dfrac{\widehat{y_i(\boldsymbol{x})} - L_i}{T_i - L_i}\right)^r & \text{if } L_i < \widehat{y_i(\boldsymbol{x})} < T_i \\ \left(\dfrac{\widehat{y_i(\boldsymbol{x})} - U_i}{T_i - U_i}\right)^r & \text{if } T_i < \widehat{y_i(\boldsymbol{x})} < U_i \\ 1 & \text{if } \widehat{y_i(\boldsymbol{x})} = T_i \end{cases} \tag{19}$$

- Individual desirability function for the smaller the better (STB) case:

$$d_i = \begin{cases} 1 & \text{if } \widehat{y_i(\boldsymbol{x})} \le L_i \\ \left(\dfrac{\widehat{y_i(\boldsymbol{x})} - U_i}{L_i - U_i} \right)^r & \text{if } L_i < \widehat{y_i(\boldsymbol{x})} < U_i \\ 0 & \text{if } \widehat{y_i(\boldsymbol{x})} \ge U_i \end{cases} \tag{20}$$

- Individual desirability function for the larger the better (LTB) case:

$$d_i = \begin{cases} 0 & \text{if } \widehat{y_i(\boldsymbol{x})} \le L_i \\ \left(\dfrac{\widehat{y_i(\boldsymbol{x})} - L_i}{U_i - L_i} \right)^r & \text{if } L_i < \widehat{y_i(\boldsymbol{x})} < U_i \\ 1 & \text{if } \widehat{y_i(\boldsymbol{x})} \ge U_i \end{cases} \tag{21}$$

The values of L_i and U_i are some acceptable lower bound and upper bound for $\widehat{y_i(\boldsymbol{x})}$, and T_i is its target value. The value of r can be chosen so that the desirability criterion is easier or more difficult to satisfy, indicating the weight of $\widehat{y_i(\boldsymbol{x})}$ in the process.

The functions in (19), (20), and (21) can be expressed in terms of $\hat{y}_{Mean}(\boldsymbol{x})$ and $\hat{y}_{Var}(\boldsymbol{x})$ in the case of a DOAD, and in terms of $E_z\left(\widehat{y(\boldsymbol{x},\boldsymbol{z})}\right)$ and $Var_z\left(\widehat{y(\boldsymbol{x},\boldsymbol{z})}\right)$ while dealing with a CAD.

4.2.2. Overall desirability function

For a n responses system, the overall performance of the system is determined by the overall desirability D, which can be expressed as the geometric mean:

$$D = \left(\prod_{i=1}^{n} d_i \right)^{1/n}. \tag{22}$$

The optimization problem to solve is then

$$\begin{cases} \text{Maximize} & D \\ \text{Subject to} & \boldsymbol{x} \in R. \end{cases} \tag{23}$$

V. ILLUSTRATIVE EXAMPLE AND RESULTS

5.1. Robust design conducted on a chemical process

5.1.1. Problem statement and design

The application is a chemical process adapted from Lawson [14]. In this process, side reactions create tars that result in lower product quality. When the level of tars produced is too high, yield decreases and further blending must be done with the finished product to decrease tar to a level that is acceptable to customers. The tars were thought to be created by reactions involving impurities in the most expensive reagent, A, and other impurities that accumulate in the recycle solvent stream. The proposed method to solving the problem was to experiment with the setting of process variables, or design parameters, to see if operating conditions could be found that were less sensitive to impurities in reagent A and the solvent stream (Lawson [15]).

The objectives of conducting a designed experiment are to diminish the proportion of impurities and to reduce the variance of the process.

The response variable is the proportion of impurities (in percentage). The factors involved in this experiment are 3 design factors and 2 noise factors. The design factors are x_1: reaction temperature, x_2: the catalyst concentration, x_3: the excess of reagent B. The noise factors are z_1: purity of reagent A, z_2: purity of the solvent stream. It is assumed that z_1 and z_2 are uncorrelated and $\sigma^2_{z_1} = \sigma^2_{z_2} = 1$ so that the variance-covariance matrix, say $\text{cov}(z) = \Omega$, is an identity matrix of dimension 2. Table 7(a) shows the coded levels of the factors and their corresponding real values.

Experiments are performed at combinations of levels of the design factors defined by a Box- Behnken design. The combinations of levels of the noise factors are arranged in a 2^2 factorial design. This means that an experiment for the 15 combinations of the control factors is realised, and each of these is repeated in each of the possible combinations of the noise factors. The results obtained from the experiment are shown in Table 7(b).

5.1.2. Methods of analysis

We recall that the aim of the experiment is to diminish the proportion of impurities. To obtain the optimal settings, we apply the optimization method based on Mean Square Error Criterion where the target value is set to $T = 0$. In addition, we apply the desirability function and the results obtained from both methods are compared.

5.1.3. Results

Case of double orthogonal array design

From the results of the experiment given in annexes (Table 7(b)), the estimated mean response and estimated variance response are determined. Terms with nonsignificant effects are not considered. The fitted models for the mean and variance are the following:

$$\hat{y}_{Mean}(\boldsymbol{x}) = 14.80 - 8.17x_1 - 9.09x_2 + 0.52x_1^2 + 8.30x_1x_2 + 5.01x_2^2$$

$$\hat{y}_{Var}(\boldsymbol{x}) = \left(\hat{y}_{SD}(\boldsymbol{x})\right)^2 = \left(3.66 - 4.44x_2 + 1.64x_3 + 2.55x_2^2 + 1.61x_3^2\right)^2.$$

- ***MSE approach***

The optimization problem to solve is:

$$\begin{cases} \text{Minimize} & \left(14.80 - 8.17x_1 - 9.09x_2 + 0.52x_1^2 + 8.30x_1x_2 + 5.01x_2^2\right)^2 + \left(3.66 - 4.44x_2 + 1.64x_3 + 2.55x_2^2 + 1.61x_3^2\right)^2 \\ \text{Subject to} & -1 \le x_1 \le 1 \\ & -1 \le x_2 \le 1 \\ & -1 \le x_3 \le 1. \end{cases}$$

We solve this optimization problem using the *nloptr* package (Ypma [16]).

- ***Desirability function approach***

The lower and upper bound for the mean response are respectively given by $L_{Mean} = \hat{y}_{Mean}(\boldsymbol{x}_{\min})$ and $U_{Mean} = \hat{y}_{Mean}(\boldsymbol{x}_{\max})$, where $\boldsymbol{x}_{\min}$ and $\boldsymbol{x}_{\max}$ are respectively solutions of the following optimization problems:

$$\begin{cases} \text{Minimize} & \hat{y}_{Mean}(\boldsymbol{x}) \\ \text{Subject to} & \boldsymbol{x} \in R \end{cases} \text{ and } \begin{cases} \text{Maximize} & \hat{y}_{Mean}(\boldsymbol{x}) \\ \text{Subject to} & \boldsymbol{x} \in R. \end{cases}$$

Similar logic leads to $L_{Var} = \hat{y}_{Var}(\boldsymbol{x}_{\min})$ and $U_{Var} = \hat{y}_{Var}(\boldsymbol{x}_{\max})$, where $\boldsymbol{x}_{\min}$ and $\boldsymbol{x}_{\max}$ are respectively solutions of the optimization problems:

$$\begin{cases} \text{Minimize} & \hat{y}_{Var}(\boldsymbol{x}) \\ \text{Subject to} & \boldsymbol{x} \in R \end{cases} \text{ and } \begin{cases} \text{Maximize} & \hat{y}_{Var}(\boldsymbol{x}) \\ \text{Subject to} & \boldsymbol{x} \in R. \end{cases}$$

Table 3 displays the values obtained.

Table 3: Lower bounds and upper bounds of individual desirability functions.

L_{Mean} :	7.12	L_{Var} :	1.72
U_{Mean} :	45.9	U_{Var} :	112.8

The individual desirability functions for the proportion of impurities and its variance are respectively

$$d\left(\hat{y}_{Mean}(\boldsymbol{x})\right) = \begin{cases} 1 & \text{if } \hat{y}_{Mean}(\boldsymbol{x}) \le 7.12 \\ \dfrac{\hat{y}_{Mean}(\boldsymbol{x}) - 45.9}{7.12 - 45.9} & \text{if } 7.12 < \hat{y}_{Mean}(\boldsymbol{x}) < 45.9 \\ 0 & \text{if } \hat{y}_{Mean}(\boldsymbol{x}) \ge 45.9 \end{cases}$$

and

$$d\left(\hat{y}_{Var}(\boldsymbol{x})\right)=\begin{cases}1 & \text{if } \hat{y}_{Var}(\boldsymbol{x})\le 1.72\\ \dfrac{\hat{y}_{Var}(\boldsymbol{x})-112.8}{1.72-112.8} & \text{if } 1.72<\hat{y}_{Var}(\boldsymbol{x})<112.8\\ 0 & \text{if } \hat{y}_{Var}(\boldsymbol{x})\ge 112.8\end{cases}$$

The overall desirability function is

$$D=\sqrt{d\left(\hat{y}_{Mean}(\boldsymbol{x})\right)\times d\left(\hat{y}_{Var}(\boldsymbol{x})\right)}.$$

The optimization problem to solve is the following:

$$\begin{cases}\text{Maximize} & D\\ \text{Subject to} & -1\le x_1\le 1\\ & -1\le x_2\le 1\\ & -1\le x_3\le 1.\end{cases}$$

We adapt the *desirability* package (Kuhn [17]) to solve the previous optimization problem.

Case of combined array design

The fitted regression model is:

$$\widehat{y(\boldsymbol{x},\boldsymbol{z})}=14.79-8.17x_1-9.09x_2+8.30x_1x_2+5.01x_2^2+3.91z_1-1.20z_2-3.30x_2z_1.$$

The mean and variance models are:

$$E_z\left(\widehat{y(\boldsymbol{x},\boldsymbol{z})}\right)=14.79-8.17x_1-9.09x_2+8.30x_1x_2+5.01x_2^2$$

$$Var_z\left(\widehat{y(\boldsymbol{x},\boldsymbol{z})}\right)=16.73-25.81x_1+7.92x_2+10.89x_1^2+10.89x_2^2.$$

- ***MSE approach***

The optimization problem to be solved is:

$$\begin{cases}\text{Minimize} & \left(14.79-8.17x_1-9.09x_2+8.30x_1x_2+5.01x_2^2\right)^2+\left(16.73-25.81x_1+7.92x_2+10.89x_1^2+10.89x_2^2\right)\\ \text{Subject to} & -1\le x_1\le 1\\ & -1\le x_2\le 1\\ & -1\le x_3\le 1.\end{cases}$$

- ***Desirability function approach***

The lower and upper bound for the mean response are respectively given by $L_{Mean}=E_z\left(\widehat{y(\boldsymbol{x}_{\min},\boldsymbol{z})}\right)$ and $U_{Mean}=E_z\left(\widehat{y(\boldsymbol{x}_{\max},\boldsymbol{z})}\right)$, where $\boldsymbol{x}_{\min}$ and $\boldsymbol{x}_{\max}$ are respectively solutions of the following optimization problems:

$$\begin{cases}\text{Minimize} & E_z\left(\widehat{y(\boldsymbol{x},\boldsymbol{z})}\right)\\ \text{Subject to} & \boldsymbol{x}\in R\end{cases} \text{ and } \begin{cases}\text{Maximize} & E_z\left(\widehat{y(\boldsymbol{x},\boldsymbol{z})}\right)\\ \text{Subject to} & \boldsymbol{x}\in R.\end{cases}$$

Similar logic leads to $L_{Var}=Var_z\left(\widehat{y(\boldsymbol{x}_{\min},\boldsymbol{z})}\right)$ and $U_{Var}=Var_z\left(\widehat{y(\boldsymbol{x}_{\max},\boldsymbol{z})}\right)$, where $\boldsymbol{x}_{\min}$ and $\boldsymbol{x}_{\max}$ are respectively solutions of the optimization problems:

$$\begin{cases}\text{Minimize} & Var_z\left(\widehat{y(\boldsymbol{x},\boldsymbol{z})}\right)\\ \text{Subject to} & \boldsymbol{x}\in R\end{cases} \text{ and } \begin{cases}\text{Maximize} & Var_z\left(\widehat{y(\boldsymbol{x},\boldsymbol{z})}\right)\\ \text{Subject to} & \boldsymbol{x}\in R.\end{cases}$$

Table 4 displays the values obtained.

Table 4: Lower bounds and upper bounds of individual desirability functions.

L_{Mean} :	6.59	L_{Var} :	0.37
U_{Mean} :	45.4	U_{Var} :	56.4

The individual desirability functions for the proportion of impurities and its variance are:

$$d\left(E_z\left(\widehat{y(\boldsymbol{x},\boldsymbol{z})}\right)\right)=\begin{cases}1 & \text{if } E_z\left(\widehat{y(\boldsymbol{x},\boldsymbol{z})}\right)\le 6.59\\ \dfrac{E_z\left(\widehat{y(\boldsymbol{x},\boldsymbol{z})}\right)-45.4}{6.59-45.4} & \text{if } 6.59<E_z\left(\widehat{y(\boldsymbol{x},\boldsymbol{z})}\right)<45.4\\ 0 & \text{if } E_z\left(\widehat{y(\boldsymbol{x},\boldsymbol{z})}\right)\ge 45.4\end{cases}$$

and

$$d\left(Var_z\left(\widehat{y(\boldsymbol{x},\boldsymbol{z})}\right)\right)=\begin{cases}1 & \text{if } Var_z\left(\widehat{y(\boldsymbol{x},\boldsymbol{z})}\right)\le 0.37\\ \dfrac{Var_z\left(\widehat{y(\boldsymbol{x},\boldsymbol{z})}\right)-56.4}{0.37-56.4} & \text{if } 0.37<Var_z\left(\widehat{y(\boldsymbol{x},\boldsymbol{z})}\right)<56.4\\ 0 & \text{if } Var_z\left(\widehat{y(\boldsymbol{x},\boldsymbol{z})}\right)\ge 56.4.\end{cases}$$

The overall desirability function is

$$D=\sqrt{d\left(E_z\left(\widehat{y(\boldsymbol{x},\boldsymbol{z})}\right)\right)\times d\left(Var_z\left(\widehat{y(\boldsymbol{x},\boldsymbol{z})}\right)\right)}.$$

The optimization problem to solve is the following:

$$\begin{cases}\text{Maximize} & D\\ \text{Subject to} & -1\le x_1\le 1\\ & -1\le x_2\le 1\\ & -1\le x_3\le 1.\end{cases}$$

The results of these optimization problems are in Table 5.

Table 5: Results.

Design	Statistic	MSE	DF
DOAD	$\boldsymbol{x}_{Opt}$	$(1,\ 0.1966171,\ -0.5093168)$	$(1,\ 0.39689872,\ -0.50935614)$
	$\hat{y}_{Mean}(\boldsymbol{x}_{Opt})$	7.19	7.63
	$\hat{y}_{Var}(\boldsymbol{x}_{Opt})$	6.09	3.54
	$\widehat{MSE}$	57.8	61.7
	Desirability	–	0.98528083
CAD	$\boldsymbol{x}_{Opt}$	$(0.99999917,\ 0.01634704,\ -0.42796747)$	$(1,\ -0.18672138,\ 0.87745330)$
	$E_z\left(\widehat{y(\boldsymbol{x}_{Opt},\boldsymbol{z})}\right)$	6.61	6.94
	$Var_z\left(\widehat{y(\boldsymbol{x}_{Opt},\boldsymbol{z})}\right)$	1.94	0.711
	$\widehat{MSE}$	45.6	48.9
	Desirability	–	0.99241997

Looking at the optimal values obtained for the proportion of impurities and its variance, it seems that the combined array design performs better than the double orthogonal array design in this chemical process.

In addition, regarding the method of analysis, if the main objective of the experiment is to reduce the proportion of impurities, the mean square error approach gives the best results. On the contrary, if the interest is the reduction of variability in the process, the desirability function seems to be the best method of analysis.

When both the reduction of proportion of impurities and reduction of variability are of utmost importance, the selection of the design to be conducted and the method of analysis to be adopted will take into account the estimated value of the corresponding quality loss function.

The quality loss function accounts for the loss in terms monetary funds incurred by the deviation from the target and high variability in the process. The concept of quality loss function is illustrated through the example presented in the following section.

5.2. Economic impact of a robust design

The chemical process is working in real conditions of $x_1 = 200$, $x_2 = 28$ and $x_3 = 17$ for the reaction temperature, catalyst concentration and excess of reagent B, respectively. A sample of 24 data is taken to verify the conditions of impurities. These refer to the mean and variance of the process. The sample values are:
23.08, 23.01, 18.11, 20.14, 47.14, 23.20, 17.12, 18.93, 21.42, 22.72, 20.36, 24.99, 32.77, 25.26, 19.50, 23.14, 23.09, 30.49, 35.72, 21.95, 26.43, 24.48, 34.21, 25.03 (Mares and Domínguez [6]).

The corresponding sample mean and sample variance are

$$\bar{y} = \frac{1}{24}\sum_{i=1}^{24} y_i = 25.1 \text{ and } s^2 = \frac{1}{23}\sum_{i=1}^{24}\left(y_i - \bar{y}\right)^2 = 45.64, \text{ respectively.}$$

Therefore, the estimate of the expected quality loss (EQL) for the quadratic loss is:

$$\widehat{Q}_{process} = k\left(\bar{y}^2 + s^2\right) = (25.1)^2 + 45.64 = 675.65,$$

where $k = 1$. This constant in practice plays a very important role because it indicates the overall cost of poor quality.
We now calculate the estimate of the expected quality loss of the optimal process conditions as given by Table 5.

The estimate of the expected quality loss of the designed experiment is defined as

$$\widehat{Q}_{doe} = \begin{cases} \left[\widehat{y}_{Mean}(\boldsymbol{x}_{Opt})\right]^2 + \widehat{y}_{Var}(\boldsymbol{x}_{Opt}) & \text{(Case of DOAD)} \\ \left[E_z\left(\widehat{y(\boldsymbol{x}_{Opt},\boldsymbol{z})}\right)\right]^2 + Var_z\left(\widehat{y(\boldsymbol{x}_{Opt},\boldsymbol{z})}\right) & \text{(Case of CAD)} \end{cases}$$

The overall difference between the estimate of the expected quality loss of the real process and the estimate of the expected quality loss of the optimal process conditions of the designed experiment, this is $\Delta = \widehat{Q}_{process} - \widehat{Q}_{doe}$, indicates the expected gain incurred by conducting the process in its optimal conditions. Table 6 presents the results.

Table 6: Economic impact of designed experiment.

Approach	Economic impact	DOAD	CAD
Mean Square Error	$\widehat{Q}_{doe}$	57.78	45.63
	Δ	617.86	630.02
Desirability Function	$\widehat{Q}_{doe}$	61.75	48.87
	Δ	613.89	626.77

We can see that the estimate of the expected quality loss of the designed experiment, $\widehat{Q}_{doe}$, ranges between 45.63 and 61.75 monetary units, while the overall difference that indicates the gain incurred by conducting the process under the optimal conditions, Δ, ranges between 613.89 and 630.02 monetary units.

VI. GENERAL CONCLUSIONS

This paper presents a methodological approach for conducting robust designs and analyzing the data obtained. An overview of two approaches used to conduct robust design and analyse the data is provided. We have presented the Taguchi approach commonly known as double orthogonal array design and the combined array design. The mean square error criterion and desirability function have been used as optimization procedures to finding the optimal conditions of the process.

To highlight the practical implementation of the double orthogonal array design and combined array design, we have given an illustrative example on a chemical process where the objectives are to diminish the proportion of impurities and to reduce the variance of the process. Looking at the optimal values obtained for the proportion of impurities and its variance, it seems that the combined array design performs better than the double orthogonal array design.

Focusing on the proportion of impurities, the mean square error approach gives the best results, and when the main interest is the reduction of variability in the process, the desirability function seems to be the best method of analysis.

When both the reduction of proportion of impurities and reduction of variability are of utmost importance, the selection of the design and the method of analysis to be used must take into account the estimated value of the quality loss function. Based upon the quality loss function, we have given an illustrative example that highlights the practical importance of conducting a designed experiment in order to reduce the overall cost of the production. In this example, the estimate of the expected quality loss of the designed experiment, $\widehat{Q}_{doe}$, ranges between 45.63 and 61.75 monetary units, while the overall difference that indicates the gain incurred by conducting the process under the optimal conditions, Δ, ranges between 613.89 and 630.02 monetary units.

REFERENCES

[1]. G. Taguchi, Introduction to Quality Engineering: Designing Quality Into Products and Processes (White Plains, NY: Krauss International Publications, 1986).

[2]. J. Harrington, The Desirability Function, Industrial Quality Control, 21, 1965, 31-45.

[3]. A. Giovagnoli and D. Romano, Robust Design via Simulation Experiments: A Modified Dual Response Surface Approach. Quality and Reliability Engineering International, 24, 2008, 401-416.

[4]. A. Bizimana, Robust Design Applied to Agro-Industrial Processes. Masters degree thesis. Center for Research in Mathematics - CIMAT, Mexico, 2010.

[5]. G. Taguchi and Y. Wu, Introduction to Off-line Quality Control (Nagoya, Japan: Central Japan Quality Control Association, 1985).

[6]. A. Mares and J. D. Domínguez, Conditional expectation and variability in the industrial problem solution, IIE Annual Conference. Proceedings, 2013, 3338-3347.

[7]. R. H. Myers and W. H. Carter, Response Surface Techniques for Dual Response Systems, Technometrics, 15, 1973, 301-317.

[8]. R. H. Myers and G. G. Vining, Combining Taguchi and Response Surface Philosophies: A Dual Response Approach, Journal of Quality Technology, 22(1), 1990, 38-45.

[9]. D. K. J. Lin and W. Tu, Dual Response Surface Optimization, Journal of Quality Technology, 27(1), 1995, 34-39.

[10]. J. Meng, C. Zhang, B. Wang and W. A. Eckerle, Integrated Robust Parameter Design and Tolerance Design, Int. J. Industrial and Systems Engineering, 5(2), 2010, 159-189.

[11]. W. Y. Fowlkes and C. M. Creveling, Engineering Methods for Robust Product Design. Using Taguchi Methods in Technology and Product Development (Addisson-Wesley Publishing Company, 1995).

[12]. O. Köksoy, Multiresponse robust design: Mean square error (mse) criteria, Applied mathematics and Computation, 175(8), 2006, 1716-1729.

[13]. G. Derringer and R. Suich, Simultaneous optimization of several response variables, Journal of Quality Technology, 12(1), 1980, 214-219.

[14]. J. S. Lawson, Design and Analysis of Experiments with SAS (Chapman & Hall, 2010).

[15]. J. S. Lawson, Improving a Chemical Process Through Use Of a Designed Experiment, Quality Engineering, 3(2), 1990, 215-235.

[16]. J. Ypma, Introduction to nloptr: an R interface to NLopt, R package version 1.0.4, 2015.

[17]. M. Kuhn, The desirability Package, R package version 1.9, 2015.

VII. ANNEXES

Table 7: Data for the chemical process problem.

(a) Coded and real levels of the factors.

	Levels		
Design factors	**-1**	**0**	**1**
x_1	180	210	240
x_2	25	30	35
x_3	12	15	18
Noise factors	**-1**	**1**	
z_1	10	20	
z_2	30	40	

(b) Experimental results of the chemical process.

Experimental run	x_1	x_2	x_3	z_1 / z_2	1 / –1	–1 / –1	–1 / 1	1 / 1	**Mean**	**Standard deviation**
1	–1	–1	0		57.81	37.29	42.87	47.07	46.26	8.68
2	1	–1	0		24.89	4.35	8.23	14.69	13.04	8.98
3	–1	1	0		13.21	9.51	10.10	11.19	11.00	1.63
4	1	1	0		13.29	9.15	10.30	11.23	10.99	1.80
5	–1	0	–1		27.71	20.24	22.28	24.23	23.62	3.18
6	1	0	–1		11.40	4.48	5.44	8.23	7.39	3.11
7	–1	0	1		30.65	18.40	20.24	24.45	23.44	5.44
8	1	0	1		14.94	2.29	4.30	8.49	7.51	5.59
9	0	–1	–1		42.68	22.42	21.64	30.30	29.26	9.76
10	0	1	–1		13.56	10.08	9.85	11.38	11.22	1.70
11	0	–1	1		50.60	13.19	18.84	30.97	28.40	16.55
12	0	1	1		15.21	7.44	9.78	11.82	11.06	3.29
13	0	0	0		19.62	12.29	13.14	14.54	14.90	3.28
14	0	0	0		20.60	11.49	12.06	13.49	14.41	4.21
15	0	0	0		20.15	12.20	14.06	13.89	15.08	3.49