Centro de Investigación en Matemáticas, A.C.

Robust Design Applied to Agro-Industrial Processes

Tesina

Que para obtener el Grado de Maestría en Ciencias con especialidad en Probabilidad y Estadística

Presenta:

BIZIMANA Alphonse

Director:

Dr. Jorge Domínguez Domínguez

Guanajuato, Gto., Junio de 2010

Esta tesina corresponde a los estudios realizados con una beca otorgada por la Secretaría de Relaciones Exteriores del Gobierno de México.

Dedication

To Flora Jeannine my dear friend and wife.

To Louange Anelyse my daughter and Lagrange Alderick my son.

Contents

Acknowledgments

Thanks must go to the Ministry of Foreign Affairs of Mexico that has awarded me a scholarship to pursue a Masters Degree in Sciences with a specialization in Probability and Statistics at the Centre of Mathematical Research- CIMAT in Guanajuato, and to pursue a course in Spanish and Mexican Culture at the National University of Mexico, Campus Taxco.

I am indebted to the Centre of Mathematical Research– CIMAT, especially to the Probability and Statistics Department, for providing a fruitful academic environment.

My special thanks to Dr. Jorge Domínguez Domínguez, who has guided me through the entire period of this work. Your help not only has brought me to accomplish this research, but also has opened my mind to the marvellous importance of statistics.

It is impossible to acknowledge all the help and academic support which have come from professors of the Probability and Statistics Department, during my stay in CIMAT. In particular, I wish to thank Dr. Víctor Manuel Rivero Mercado, Dr. Miguel Nakamura Savoy and Dr. Eloísa Díaz Francés Murguía. Your advice will remain stamped in my memory. I continue to be grateful to Dr. Enrique Raúl Villa Diharce, Dr. José Miguel Ponciano Castellanos, Dr. José Alfredo López Mimbela, Dr. José Andrés Christen Gracia, Dr. Joaquín Ortega Sánchez, Dr. Rogelio Ramos Quiroga, Dr. Víctor Manuel Pérez Abreu Carrión, and Dr. Daniel Hernández Hernández, for the knowledge I acquired from you during my academic training. I owe thanks to Stephanie Dunbar for linguistic support for this thesis.

Finally, but from the bottom of my heart, I send my warm thanks to my wife UWIMABERA Flora Jeannine, to my daughter IHIRWE Louange Anelyse, and to my son BIZIMANA Lagrange Alderick. You are as much as part of this work for having allowed me to leave you for so long.

Introducción

El uso de diseño de experimentos desempeña un papel importante en varias áreas tales como Economía, Agro-industria, y Medicina.

Los ingenieros y científicos se han ido capacitando en el conocimiento y aplicación del diseño de experimentos, y a partir de ahí han surgido nuevas áreas de aplicación. Entre ellas el diseño robusto. La metodología de diseño robusto es un esfuerzo sistemático para alcanzar insensibilidad a los factores de ruido.

El supuesto es que hay dos tipos de factores que afectan a la característica de calidad. Éstos se dividen en factores de control y factores no controlables o difíciles de controlar. Estos factores se denominan respectivamente factores de diseño y factores de ruido.

Los factores de ruido se dividen en dos categorías: factores de ruido externos y factores de ruido internos. Los factores de ruido externos son fuentes de variabilidad que vienen de fuera del sistema. Ejemplos de los factores de ruido externos son factores ambientales tales como la temperatura ambiente, la presión ambiente y la humedad. Los factores de ruido internos tienen esencialmente su origen en las variaciones de los factores de control. Por ejemplo, el ruido interno incluye las desviaciones de los valores objetivos en los factores de control causadas por la manufactura, ensamble, y deterioro.

Cuando se diseña un experimento, con frecuencia es imposible o costoso controlar o eliminar la variación debida a los factores de ruido externos. Sin embargo, el experimentador tiene algún control en la determinación de los niveles de los factores de ruido internos durante el diseño. La meta del diseño robusto es permitir al experimentador elegir los niveles de los factores de control que optimizan la respuesta de interés, de tal manera que la variación causada por los factores de ruido sea mínima [32].

El diseño robusto se compone principalmente de tres etapas: el diseño robusto de sistema, el diseño robusto de parámetro y el diseño robusto de tolerancia [18]. El diseño de sistema consiste en usar la física, las matemáticas, la experiencia y el conocimiento adquirido en un campo específico, para desarrollar y seleccionar las condiciones del diseño más apropiadas. Una vez que la configuración de un sistema se establece, se determinan los ajustes nominales y las tolerancias de las variables de diseño. El objetivo del diseño robusto de parámetro es encontrar los ajustes óptimos de los niveles de factores de control de manera que el sistema sea insensible o menos sensible a los factores de ruido. El diseño robusto de tolerancia es un proceso de equilibrio. Dicho de otro modo, el diseño de tolerancia busca encontrar los ajustes óptimos de las tolerancias de los factores de control, tal que el costo total del sistema sea mínimo [18].

La formalización del diseño robusto fue iniciada por Genichi Taguchi. Él introdujo el enfoque llamado diseño robusto de parámetro. Este enfoque se basa en clasificar los factores en factores de control y de ruido, y luego encontrar los ajustes para los factores de control que reducen al mínimo valor la variabilidad transmitida a la respuesta para los factores de ruido.

Taguchi propuso el uso de un producto cruzado de dos diseños experimentales, un diseño interno que contiene los factores de diseño y un diseño externo que contiene los factores de ruido. Este producto cruzado es llamado diseño doble arreglo ortogonal [10]. Las métricas usadas por Taguchi para evaluar la robustez de un diseño son la Función Cuadrática de Pérdida y la Razón Señal a Ruido.

La estrategia experimental y el diseño recomendados por Taguchi han sido criticados por varios autores. Las críticas principales son que el diseño doble arreglo ortogonal implica una gran cantidad de tratamientos y no considera interacciones entre factores de diseño y de ruido. Más detalle sobre la discusión y las críticas del enfoque de Taguchi se puede encontrar en [22] y [33]. Sin embrago, Soren Bisgaard en el 2000, abrió la discusión y la investigación sobre la interacción entre los factores de control y ruido [30].

Como alternativa al método propuesto por Taguchi, otros estadísticos (como Wu et al.[16]) adoptaron el uso del diseño arreglo combinado de factores de diseño y de ruido. Este método permite el análisis de interacciones entre ambos tipos de factores.

Los diseños comúnmente usados en conducir el diseño doble arreglo ortogonal o el diseño arreglo combinado son diseños factoriales y diseños factoriales fraccionados [03]. E. P. Box y Jones Jones [12] introdujeron el uso del diseño de parcelas divididas para la experimentación robusta.

La metodología de superficie de respuesta es una herramienta excelente para analizar los datos obtenidos del diseño arreglo combinado. Esta metodología es una colección de técnicas matemáticas y estadísticas que son útiles para modelar y analizar los problemas en los cuales una respuesta de interés depende de varias variables, y el objetivo es optimizar esta respuesta.

La metodología estadística esencial para el diseño robusto que ha sido extensamente aceptada, es la metodología de superficie de respuesta dual. Se estiman dos superficies, una para la media y otra para la varianza [01]. Se ha introducido una modificación. Esta modificación se relaciona con la división de factores de ruido dentro de dos grupos. El primer grupo consiste en factores aleatorios y el segundo se compone de los factores de ruido para los cuales los niveles son fijos. Se simula el experimento con la computadora. De los datos obtenidos mediante simulación, se calculan la media y la varianza.

Los valores nominales de los factores que se obtuvieron en el experimento con el tiempo sufren desviaciones, eso da lugar a evaluar las tolerancias. Se presenta la necesidad de aplicar el método integrado de diseño de parámetro y de diseño de tolerancia. El objetivo adicional de minimizar el costo debido a la introducción de las tolerancias en el proceso compite con el objetivo de minimizar la varianza [01].

El tema de la superficie de respuesta dual modificada es la parte sustantiva de esta tesina. Además se complementa un enfoque integral al diseño de parámetro y tolerancia que son las partes básicas del diseño robusto.

La tesina se subdivide en cinco capítulos:

El Capítulo 1 presenta el diseño doble arreglo ortogonal. Describe la estructura experimental y las métricas usadas para evaluar la robustez. Se da un ejemplo ilustrativo. La finalidad de este capítulo es dar el panorama general en el que se basan los siguientes dos capítulos. Este capítulo tiene como componente resaltar el impacto económico que tiene el realizar un experimento en la industria. Este se mide aplicando la función de pérdida y las señales a ruido. Lo que permite resaltar las bondades del planteamiento propuesto por Taguchi.

El Capítulo 2 describe el diseño arreglo combinado, con énfasis en la superficie de respuesta dual y superficie de respuesta dual modificada. Se presenta un ejemplo ilustrativo. El resultado de este capítulo permitirá tener una generalización de la superficie de respuesta dual. Este contiene una interesante aplicación de la esperanza condicional.

El Capítulo 3 presenta el método integrado de diseño de parámetro y de diseño de tolerancia. El método para minimizar el costo total del proceso está dado. Este capítulo proporciona un proceso metodológico para integrar el diseño de parámetro y tolerancia. Este se ilustra con un ejemplo.

El Capítulo 4 explica la aplicación del diseño a parcelas divididas en la experimentación robusta. Tres ordenamientos de los factores de diseño con respecto a los factores de ruido se presentan. Se da un ejemplo ilustrativo en el cual los factores de ruido están colocados en toda-la-parcela y los factores de diseño en sub-parcelas.

El Capítulo 5 presenta las conclusiones generales de este trabajo.

Introduction

The use of designed experiments plays an important role in several fields such as Economy, Agro-industry, and Medicine.

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 in 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. These are the control factors and the uncontrollable or difficult to control factors. They are respectively referred to as design factors and noise factors.

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 manufacturing, assembly, and deterioration.

While designing an experiment, it is frequently impossible or very difficult or expensive to control or eliminate sources of variation due to external noise factors. However, the experimenter has some control on setting the levels of internal noise factors during the design. The goal of robust design is to enable the experimenter to choose the levels of the control factors that optimize a defined response while minimizing the variation imposed on the process via the noise factors [32].

Robust design is mainly composed of three stages: robust system design, robust parameter design and robust tolerance design [18]. 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 to find the optimal settings of tolerances of the control factors so that the total cost of the system is minimal [18].

The formalization of robust design was initiated by Genichi Taguchi. He introduced an approach referred to as robust parameter design. His approach is based on classifying the factors as either controllable or noise factors, and then finding the settings for the controllable factors that minimize the variability transmitted to the response from the noise factors.

Taguchi proposed the use of a cross-product of two experimental designs, known as product array design or double orthogonal array design [10]. This consists of an inner array containing the design factors and an outer array containing the noise factors. For each combination of design factors, the same array of noise factors is run. The metrics used by Taguchi for evaluating the robustness of a design are the Quadratic Loss Function and the Signal-to-Noise Ratio.

Experimental strategy and design advocated by Taguchi have been criticized by various authors. The main criticism is that the double array design involves a large amount of runs and does not consider interactions between design and noise factors. Details of discussion and criticism of Taguchi´s approach can be found in [22] and [33]. However, in 2000, Soren Bisgaard opened the discussion and the investigation about interactions between the design and noise factors [30].

As an alternative to the method proposed by Taguchi, other statisticians (such as Wu et al. [16]) have adopted the use of combined array design which contains controllable and noise factors. This method permits the analysis of interactions between both kinds of factors.

The designs commonly used in conducting double orthogonal array design or combined array design are factorial and fractional factorial designs [03]. George E. P. Box and Jones Jones [12] have introduced the use of split- plot design for robust experimentation.

Response surface methodology is an excellent tool for analysing the data obtained from combined array design. This is a collection of mathematical and statistical techniques that are useful for modeling and analysing problems in which a response of interest depends on several variables, and the objective is to optimize this response.

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 one for the variance of the quality characteristic [01]. A modification has been introduced. This modification is related to the division of the noise factors within two groups. The first group consists of random factors and the second is composed of the noise factors for which the levels are fixed. The experiment with the computer is simulated. From the data obtained by simulation, the mean and the variance are calculated.

When the design scope is extended to the specification of allowable deviations of parameters from the nominal settings (tolerances), the integrated parameter and tolerance design problems arise. The additional objective of minimizing the production costs needed to fulfil tolerance specifications will compete with the minimum variance objective [01].

The topic of modified dual response surface is the substantive part of this thesis. In addition, the thesis is completed by an integrated approach of parameter design and tolerance design, which are the basic parts of robust design.

The thesis is subdivided into five chapters:

Chapter 1 presents Double Orthogonal Array Design. It describes the experimental structure and the metrics used to evaluate the robustness. An illustrative example is given. The purpose of this chapter is to give a general view on which the following two chapters are based. The chapter emphasizes the economic impact of conducting designed experiments in the industry. The impact is measured by applying the Quality Loss Function and the Signal-to-Noise Ratio. These highlight the aspects of robustness proposed by Taguchi.

Chapter 2 describes Combined Array Design. Emphasis is given to Dual Response Surface and Modified Dual Response Surface with an illustrative example. The results of this chapter permit a generalization of the Dual Response Surface. This contains an interesting application of the conditional expectation.

Chapter 3 presents the integrated method of Parameter design and Tolerance design. The method for minimizing the overall cost of the process is given. This chapter provides a methodological process of applying the integrated method of Parameter design and Tolerance design. The method is illustrated by an example.

Chapter 4 explains the application of split-plot designs in robust experimentation. Three arrangements of design and noise factors are presented. An illustrative example is given. In this example, the whole plots are formed by the noise factors, and the design factors are in sub plots.

Chapter 5 presents the general conclusions of this work.

Chapter 1: Double Orthogonal Array Design

1.1. Overview of Double Orthogonal Array Design

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

1.2. Experimental structure

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

		z_1	z_{11}	$\cdots$	z_{r1}			
		$\vdots$	$\vdots$	$\vdots$	$\vdots$			
		z_q	z_{1q}	$\cdots$	z_{rq}			
x_1	$\cdots$	x_p	**Observations**			$\overline{y}$	s^2	SNR
x_{11}	$\cdots$	x_{1p}	y_{11}	$\cdots$	y_{1r}	$\overline{y}_1$	s_1^2	SNR_1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_{n1}	$\cdots$	x_{np}	y_{n1}	$\cdots$	y_{nr}	$\overline{y}_n$	s_n^2	SNR_n

Figure 1.1: Experimental structure of Double Orthogonal Array Design.

Data analysis

The data generated by the double orthogonal array design permit modelling the mean and the standard deviation or the variance of the characteristic of interest. The metrics proposed by Taguchi for evaluating the robustness are the Quadratic Loss Function (QLF) and the Signal-to-Noise Ratio (SNR). However, the use of the SNR has drawn much criticism [33].

The primary goal of the Taguchi philosophy, to obtain a target condition on the mean while minimizing the variance, can be achieved within response surface methodology framework. By combining Taguchi and response surface philosophies, a dual response surface approach will be applied. This consists of fitting regression models for the mean and the variance. The dual response surface provides a more rigorous method for achieving a target for the mean, while also achieving a target for the variance [25]. The optimization based upon the criteria of mean square error gives a fairly general method to solve the dual response surface problem [05].

1.2.1. Data analysis considering 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 [18]. In this thesis the QLF is used to evaluate the economic impact of conducting an experiment on the process.

The QLF is given by the expression

$$L(y) = k(y-M)^2 \quad . \tag{1.1}$$

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

The expected quality loss is $Q = E\left[L(y)\right] = k\,E(y-M)^2 = k\left[\left(E(y)-M\right)^2 + Var(y)\right]$.

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

$Q = k\left[(\mu - M)^2 + \sigma^2\right]$.

Then the estimate of the expected quality loss is:

$$\hat{Q} = k\left[\left(\hat{\mu} - M\right)^2 + \widehat{\sigma^2}\right]; \tag{1.2}$$

where $\hat{\mu} = \bar{y}$ 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 functional limits are determined by $M \pm \Delta_0$. 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 $M \pm \Delta_0$, that is $L(y) = A_0$ at $y = M \pm \Delta_0$. Substituting the functional limits $M \pm \Delta_0$ and the value of the quality loss into Equation (1.1), the quality loss coefficient is found to be

$$k = \frac{A_0}{(\Delta_0)^2}. \tag{1.3}$$

1.2.2. Types of Quadratic Loss Function

While conducting an experiment, the designer is interested in reaching the target or minimizing or maximizing the value of the quality characteristic. These three cases of quality characteristic are referred to as Nominal the best, Smaller the better and Larger the better.

Let $y^T = (y_1, \cdots, y_n)$ where y is the quality characteristic of a product or process. Table 1.1 shows the types of QLF and the average quadratic loss functions corresponding to each kind of quality characteristic of interest. More details of this section can be found in [35] and [37].

Table 1.1: Types of Quadratic Loss Function.

Type	Quadratic Loss	Expected Quadratic Loss	Estimate of Expected Quadratic Loss
The Nominal the Best	$L(y) = A(y-M)^2$	$Q = A\left((\mu - M)^2 + \sigma^2\right)$	$\widehat{Q} = A\left((\overline{y} - M)^2 + s^2\right)$
The Smaller the Better	$L(y) = Ay^2$	$Q = A\left(\mu^2 + \sigma^2\right)$	$\widehat{Q} = A\left((\overline{y})^2 + s^2\right)$
The Larger the Better	$L(y) = \dfrac{A'}{y^2}$	$Q \doteq \dfrac{A'}{\mu^2}\left[1 + \dfrac{3\sigma^2}{\mu^2}\right]$	$\widehat{Q} \doteq \dfrac{A'}{(\overline{y})^2}\left[1 + \dfrac{3s^2}{(\overline{y})^2}\right]$

In this table, $A = \dfrac{A_0}{(\Delta_0)^2}$ and $A' = A_0 (\Delta_0)^2$.

1.2.3. Limitations of Quadratic Loss Function

Taguchi´s QLF has its limitations. It does not apply when there is a range $\pm\delta$ around the target M where customers cannot tell the difference [18]. The Loss Function applied in this case is:

$$L(y) = \begin{cases} 0 & \text{for } -\delta \leq y - M \leq \delta \\ k(y-M)^2 - k\delta^2 & \text{for } -\infty < y - M < -\delta \text{ or } \delta < y - M < \infty \end{cases}. \tag{1.4}$$

Furthermore, Taguchi´s QLF is not convenient when the loss is not symmetric around the target M . Young J. K and Byung R. C. [36] propose the following Loss Function:

$$L(y)=\begin{cases} k_1(y-M)^2 & \text{if } y \geq M \\ k_2(y-M)^2 & \text{otherwise} \end{cases} . \tag{1.5}$$

The constants k_1 and k_2 are positive loss coefficients. More details about the Asymmetric QLF can be found in [36].

1.2.4. Data analysis considering the Signal-to-Noise Ratio

Although the QLF is a metric of robustness, it has deficiencies as mentioned in the preceding sections. Taguchi suggests that the response values at each inner array design point be summarized by a performance criterion called Signal- to-Noise Ratio (SNR). He determined various forms of the SNR. In this thesis we describe three of those functions as they appear in the majority of published works. The SNR is a statistic that estimates the effect of noise factors on the quality characteristic. Data analysis considering the SNR permits identifying the factors which are important in the process, and their corresponding levels. The SNR we consider are determined in terms of the decimal logarithm of the mean square deviation of the quality characteristic from the target. There are three types of SNR depending on the desired performance response [35].

The smaller the better: This SNR is used when the experimenter is interested in minimizing the system response.

It is calculated as follows: $SNR_S = -10\log\left(\frac{1}{n}\sum_{i=1}^{n} y_i^2\right)$

The larger the better: It is used when the experimenter is interested in maximizing the response and it is given by the expression $SNR_L = -10\log\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{y_i^2}\right)$.

The nominal the best: This SNR is used when the experimenter needs the response to attain a certain target value. It is given by $SNR_T = 10\log\left(\frac{\bar{y}^2}{s^2}\right)$.

The preferred parameter settings are determined through analysis of the SNR, where the levels of the factors that maximize the appropriate SNR are optimal. More detailed literature on the SNR is in reference [13].

1.2.5. Agro-industrial processes

Agro-industry includes various areas such as food technology, food processing and agricultural materials among others [38].

Food technology deals with sources of food, raw material sorting, postharvest transformation, principles of food preservation and processing, and roles of microorganism.

Food processing deals with preparation of raw materials for food processing, fermentation, low and high temperature processing.

Agricultural materials include products such as fertilizers, pesticides and other materials used in different agricultural activities.

Robust design in agro-industrial processes aims at maintaining or improving the quality characteristics of the response of interest by conducting designed experiments.

The applications given in this thesis are oriented to chemical processes. Their implementation in agro-industrial processes is immediate. For instance, the chemical process for which the aim is to reduce the number of impurities is applied in production of fertilizers and pesticides.

1.2.6. Illustration of QLF and SNR

Let y be the quality characteristic of a chemical process. Assume that the deviation of $\Delta_0 = 2$ units from the target $M = 9$ incurs a loss of 100 monetary units. This means that if $y = M \pm \Delta_0$, the corresponding loss is $L(y) = 100$. Introducing $y = M \pm \Delta_0$ and $L(y) = 100$ into Equation(1.1) leads to $k = \frac{100}{\Delta_0^2} = \frac{100}{2^2} = 25$.

The QLF of the process is $L(y) = 25(y-9)^2$. This expression permits obtaining the loss suffered by process for any value of the quality characteristic y. For instance, the loss caused by the value of quality $y = 12$ increases to $L(12) = 25(12-9)^2 = 225$ monetary units.

Now let us consider a series of 15 random observations of the quality characteristic of the same chemical process:

$$y = (17.65 \quad 11.77 \quad 10.73 \quad 18.31 \quad 18.28 \quad 20.04 \quad 16.29 \quad 19.68 \quad 15.37 \quad 16.32 \quad 21.43 \quad 19.45 \quad 17.42 \quad 13.19 \quad 20.17).$$

By equation (1.2), the expected loss is given by $Q = 25\left((\mu-9)^2 + \sigma^2\right)$.

The corresponding estimate of expected loss is $\hat{Q} = 25\left(\left(\hat{\mu}-9\right)^2 + \widehat{\sigma^2}\right)$ where $\hat{\mu} = \bar{y} = \frac{1}{15}\sum_{i=1}^{15} y_i$ and $\widehat{\sigma^2} = s^2 = \frac{1}{14}\sum_{i=1}^{15}\left(y_i - \bar{y}\right)^2$.

Numerically: $\bar{y} = 17.07$, $s^2 = 10.05$ and $\hat{Q} = 25\left((17.07-9)^2 + 10.05\right) = 1879.4$ monetary units. This is the value of the expected quality loss for each unit. The total quality loss is calculated by multiplying the quality loss for each unit by the number of units produced. For instance, the expected quality loss for 1000 units produced is 1,879,400 monetary units.

The SNR is $SNR_T = 10\log\left(\frac{\bar{y}^2}{s^2}\right) = 10\log(28.9935) = 14.6230$.

If we consider the problem of minimizing the number of impurities, the expected loss for each unit produced becomes $\hat{Q} = 25\left((17.07)^2 + 10.05\right) = 7\,476.25$ monetary units. The corresponding SNR is $SNR_S = -10\log\left(\frac{1}{n}\sum_{i=1}^{n} y_i^2\right) = -24.78$

This illustration highlights real operating conditions of the chemical process. The improvement of this process is reached by conducting a designed experiment. This point will be looked at in the following section through a more detailed application of double orthogonal array design.

1.3. Application of Double Orthogonal Array Design

The application is a chemical process adopted from John S. Lawson [19]. In general, the constructed data would only be appropriate for the model that reflects the way in which the experiment is carried out. However, the same application will be used in Chapter 2 by using a double orthogonal array design, and in Chapter 4, as split-plot design. The results will be compared in order to highlight the efficiency of each experimental structure and the corresponding data analysis.

Motivation

One way for chemical processing companies to reduce variability in their products is to insist on higher quality or more uniform raw material, and to tighten control on other process operating conditions. In order to produce more uniform end-products, chemical processing companies are demanding higher quality, i.e. more uniformity, from suppliers of bulk industrial grade chemicals that once competed in a commodity type market. Companies that are demanding reduced variability in raw materials may have to bear the suppliers´ cost for adding and maintaining new process steps or equipment, and costs for keeping higher inventories.

Variability in the quality of product from a chemical process can also be caused by changes in some operating conditions which may be difficult to monitor or control. Maintaining tight control of such processing conditions may require additional expense. Thus, achievement of both high quality and low cost may be contradictory goals for some chemical processing companies.

A possible solution to the problem of increasing the quality of chemical products without increasing material or processing costs, involves experimentation with process variables which can be changed easily and inexpensively. Chemical process companies can seek process conditions that will provide the best quality product, regardless of fluctuations in raw materials and process variables that are difficult to monitor and control. This consists of identifying the process operating conditions that result in optimum performance while simultaneously minimizing the effect of variables which are not directly under the processing company´s control.

Formulation of the problem

The chemical process generates impurities. As a result, the product obtained has low quality. The objectives of conducting a designed experiment are:

- To diminish the number of impurities;
- To reduce the variance of the process;
- To reduce the cost of the process.

The response variable is the number of impurities (in percentage). 3 design factors and 2 noise factors are involved in this experiment. 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. Annex 1.1 shows the coded levels of the factors and their corresponding real values.

In this experiment, the goal of the parameter design is to find a combination of the temperature, catalyst concentration, and the excess of reagent B that gives good results. To do this, 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. A total of 60 runs is realized, and the results obtained are shown in Annex 1.2. As the objective is to diminish the number of impurities, the SNR to be used is $SNR_S = -10\log\left(\frac{1}{n}\sum_{i=1}^{n} y_i^2\right)$.

1.3.1. Statistical results

The idea is to fit the second order regression models of the form $y = \beta_0 + x^T\beta + x^T\mathrm{B}\,x + \varepsilon$. In this model, x is the vector of control factors, β_0 the intercept, β is a vector of coefficients of 1st order control factors, B is a matrix of coefficients of 2nd order terms of control factors and their interactions, ε is a vector of residual errors of the regression model. The residual errors are assumed to be $N\left(0,\sigma^2\right)$. The response y is the mean, the standard deviation, the variance or the SNR. After the estimate of the regression model $\hat{y} = \hat{\beta}_0 + x^T\hat{\beta} + x^T\hat{\mathrm{B}}\,x$ is obtained, the optimal setting is calculated by solving the following optimization problem:

$$\begin{cases} \text{Optimize } \hat{y} \\ x \in R \end{cases},$$

where $\hat{y}$ represents the estimated response surface for the standard deviation and the SNR. The problem becomes a minimization problem in the case of $\hat{y}_{SD}$. It is a maximization problem for $\hat{y}_{SNR}$.

As the aim of the experiment is to diminish the number of impurities, it is convenient to apply the optimization method proposed by Dennis K. J. Lin and Wanzhu Tu [05]. In [18], the method is referred to as Squared Error Criterion. The method consists of three steps:

Step 1: Find the fitted models $\hat{y}_{mean}$ and $\hat{y}_{Var}$.

Step 2: Find x_{Opt} such that the mean square error given by the expression $MSE = \left[\hat{y}_{mean}\left(x_{Opt}\right) - T\right]^2 + \hat{y}_{Var}\left(x_{Opt}\right)$ is minimized. x_{Opt} is obtained by minimizing MSE subject to $x \in R$. In this case, the target $T = 0$.

Step 3: Evaluate $\hat{y}_{mean}\left(x_{Opt}\right)$ and $\hat{y}_{Var}\left(x_{Opt}\right)$.

1.3.2. Analysis based on full design

Coefficients of the regression models

Table 1.2 gives the coefficients of the regression models for the variables mean, standard deviation, variance and SNR.

Table 1.2: Coefficients of the regression models.

	Intercept	Linear			Quadratic			Interaction		
	β_0	β_1	β_2	β_3	β_{11}	β_{22}	β_{33}	β_{12}	β_{13}	β_{23}
Mean	14.7967	-8.1738	-9.0862	-0.135	0.5154	5.0104	0.1780	8.3025	0.075	0.175
Stand. Dev.	3.66	0.0625	-4.45	1.64	-0.9475	2.5475	1.6175	-0.045	0.055	-1.3
Variance	13.55	0.87	-63.23	28.46	-24.39	51.25	30.98	-1.21	0.53	-42.69
SNR	-23.5988	3.4936	4.0921	-0.2018	0.6850	-1.7983	-0.0916	-2.4485	-0.2645	-0.2645

Regression models

The results of analysis of variance for the mean, standard deviation, variance and SNR are shown in Annexes 1.3, 1.4, 1.5 and 1.6. The corresponding estimates of the regression models considering the significant effects are the following:

- Estimated mean response surface:

$$\hat{y}_{mean} = 14.7967 - 8.1738x_1 - 9.0862x_2 + 0.5154x_1^2 + 5.0104x_2^2 + 8.3025x_1x_2 .$$

- Estimated standard deviation response surface:

$$\hat{y}_{SD} = 3.66 - 4.45x_2 + 1.64x_3 + 2.5475x_2^2 + 1.6175x_3^2 - 1.3x_2x_3 .$$

- Estimated variance response surface:

$$\hat{y}_{Var} = 13.55 + 0.87x_1 - 63.23x_2 + 28.46x_3 - 24.39x_1^2 - 1.21x_1x_2 + 0.53x_1x_3 + 51.25x_2^2 - 42.69x_2x_3 + 30.98x_3^2 .$$

- Estimated SNR response surface:

$$\hat{y}_{SNR} = -23.5988 + 3.4936x_1 + 4.0921x_2 - 2.4485x_1x_2 - 1.7983x_2^2 .$$

Optimal values for the standard deviation and the SNR

The optimization problems to be solved in order to obtain the minimal standard deviation, the maximal SNR and the corresponding optimal settings are:

$$\begin{cases} \min\left\{ 3.66 - 4.45x_2 + 1.64x_3 + 2.5475x_2^2 + 1.6175x_3^2 - 1.3x_2x_3 \right\} \\ -1 \le x_2x_3 \le 1 \end{cases}$$

and

$$\begin{cases} \max\left\{ -23.5988 + 3.4936x_1 + 4.0921x_2 - 2.4485x_1x_2 - 1.7983x_2^2 \right\} \\ -1 \le x_1x_2 \le 1 \end{cases} .$$

Table 1.3 gives the optimal values for the standard deviation and SNR. It also shows the mean square error.

Table 1.3: Optimal values for the standard deviation and the SNR.

	Optimum	Combinations			MSE
		x_1	x_2	x_3	
Stand. Dev.	0.71	$\forall x_1 \in R$	0.82	-0.16	1.669
SNR	-19.73	1	0.93	$\forall x_3 \in R$	1.901

Optimal values for the mean and the variance

The optimization procedure is based on the mean square error method.
The optimization problem to be solved is:

$$\begin{cases} \min \quad \left\{ \left(14.7967 - 8.1738x_1 - 9.0862x_2 + 0.5154x_1^2 + 5.0104x_2^2 + 8.3025x_1x_2\right)^2 + \right. \\ \qquad \left. + \left(13.55 + 0.87x_1 - 63.23x_2 + 28.46x_3 - 24.39x_1^2 - 1.21x_1x_2 + 0.53x_1x_3 + 51.25x_2^2 - 42.69x_2x_3 + 30.98x_3^2\right) \right\}. \\ -1 \le x_1x_2x_3 \le 0 \end{cases}$$

Table 1.4 gives optimal settings, optimal values for the mean and variance models and the mean square error.

Table 1.4: Optimal values for the mean and the variance.

Optimal setting (x_1, x_2, x_3)	$\hat{y}_{mean}(x_{Opt})$	$\hat{y}_{Var}(x_{Opt})$	MSE
$(0.0, -0.46)$	14.80	7.01	225.95

In sub-section 1.3.5 we calculate the estimate of the expected quality loss using the formula $\widehat{Q} = 25\left(\left(\widehat{\mu}\right)^2 + \widehat{\sigma^2}\right)$. This is the case in which the objective is to obtain the smallest number of impurities. The optimal values obtained in conducting the experimental design are $\widehat{\mu} = 14.80$ and $\widehat{\sigma^2} = 7.01$. The corresponding expected quality loss function is $\widehat{Q} = 25\left((14.80)^2 + 7.01\right) = 5651.3$ monetary units.

This value shows that the process may be improved by applying optimal values obtained from the designed experiment. In fact, the reduction of the quality loss is $7476.25 - 5651.3 = 1824.9$ monetary units for each unit of chemical product. The values of the SNR obtained considering the 15 random observations of the quality characteristic and the designed experiment are respectively -24.78 and -19.73.

Interpretation

Considering the analysis based on full design data, optimal values for the standard deviation and the SNR are respectively 0.71 and -19.73. The optimal values for the mean and the variance are calculated by the method of mean square error. The pair of optimal values for the mean and the variance is $(14.80, 7.01)$.

1.3.3. Analysis based on half fractional design

When the number of factors in factorial design increases, the number of runs required for a complete replicate of the design rapidly grows. For economic reasons, fractional factorial designs are commonly used. These designs consist of a fraction of full factorial designs. For instance, the number of runs in this chemical process application is $15 \times 2^2 = 60$. The same analysis has been conducted considering half fractional design. The two fractioned groups of 30 runs for each are given in Annex 1. 7. The corresponding results for the mean, the standard deviation, the variance and the SNR are shown in Annex 1.8.

Coefficients of the regression models

Tables 1.5 and 1.6 give respectively the coefficients of the regression models for the 1st group and the 2nd group.

Table 1.5: Coefficients of the regression models for the 1st group.

	Intercept	Linear	effects		Quadratic	effects			Interaction	
	β_0	β_1	β_2	β_3	β_{11}	β_{22}	β_{33}	β_{12}	β_{13}	β_{23}
Mean	15.2917	-11.7362	- 11.9381	0.3569	-1.8233	6.6954	1.9754	10.9187	-0.8112	-1.4125
Stand. Dev.	2.8072	-1.2763	-3.6142	1.1605	-1.7695	2.1072	1.1985	1.0589	-0.2952	-0.9086
Variance	10.29	-9.29	-41.13	16.68	-27.49	33.58	23.55	12.19	-2.90	-28.33
SNR	-23.6434	6.5660	4.2935	0.3452	2.8686	-2.3092	0.3099	-4.0217	0.9387	0.5696

Table 1.6: Coefficients of the regression models for the 2nd group.

	Intercept	Linear effects			Quadratic effects			Interaction		
	β_0	β_1	β_2	β_3	β_{11}	β_{22}	β_{33}	β_{12}	β_{13}	β_{23}
Mean	14.2967	-4.6106	-6.2331	-0.6262	2.8534	3.3335	-1.6227	5.6862	0.96	1.765
Stand. Dev.	2.3146	1.0456	-1.2365	0.8096	0.6043	0.3391	-0.5324	-0.5568	0.6152	-0.6470
Variance	9.12	7.56	-9.18	4.30	4.83	3.52	-6.55	-8.63	4.04	-3.07
SNR	-23.0984	1.8192	2.8881	0.1241	-0.7176	-1.1643	0.3327	-1.9597	-0.6434	-0.8429

Regression models

The results of analysis of variance for the mean, the standard deviation, the variance and the SNR are shown by Annexes 1.9, 1.10, 1.11 and 1.12. The corresponding estimates of the regression models considering the significant effects are the following:

For the 1st group:

- Estimated mean response surface:

$$\hat{y}_{mean} = 15.2917 - 11.7362x_1 - 11.9381x_2 + 6.6954x_2^2 + 10.9187x_1x_2 \,.$$

- Estimated standard deviation response surface:

$$\hat{y}_{SD} = 2.8072 - 1.2763x_1 - 3.6142x_2 + 1.1605x_3 - 1.7695x_1^2 + 1.0589x_1x_2 + 2.1072x_2^2 - 0.9086x_2x_3 + 1.1985x_3^2 \,.$$

- Estimated variance response surface:

$$\hat{y}_{Var} = 10.89 - 9.29x_1 - 41.13x_2 + 16.68x_3 - 27.49x_1^2 + 12.19x_1x_2 - 2.90x_1x_3 + 33.58x_2^2 - 28.33x_2x_3 + 23.55x_3^2 \,.$$

- Estimated SNR response surface:

$$\hat{y}_{SNR} = -23.6434 + 6.5660x_1 + 4.2935x_2 - 4.0217x_1x_2 + 2.8686x_1^2 - 2.3092x_2^2 \,.$$

For the 2nd group:

- Estimated mean response surface:

$$\hat{y}_{mean} = 14.2967 - 4.6106x_1 - 6.2331x_2 + 2.8535x_1^2 + 3.3335x_2^2 + 5.6862x_1x_2 \,.$$

- Estimated standard deviation response surface:

$$\hat{y}_{SD} = 2.3146 + 1.0456x_1 - 1.2365x_2 + 0.8096x_3 + 0.6043x_1^2 - 0.5568x_1x_2 + 0.6152x_1x_3 + 0.3391x_2^2 - 0.6470x_2x_3 - 0.5324x_3^2 \,.$$

- Estimated variance response surface:

$$\hat{y}_{Var} = 9.12 + 7.56x_1 - 9.18x_2 + 4.30x_3 + 4.83x_1^2 - 8.63x_1x_2 + 4.04x_1x_3 + 3.52x_2^2 - 3.07x_2x_3 - 6.55x_3^2 \,.$$

- Estimated SNR response surface:

$$\hat{y}_{SNR} = -23.0984 + 1.8192x_1 + 2.8881x_2 - 0.7176x_1^2 - 1.9597x_1x_2 - 0.6434x_1x_3 - 1.1643x_2^2 - 0.8429x_2x_3 \,.$$

Optimal values for the standard deviation and the SNR

The optimization problems to be solved in order to obtain the minimal standard deviation, the maximal SNR and the corresponding optimal settings are:

For the 1st group:

$$\begin{cases} \min\left\{ 2.8072-1.2763x_1-3.6142x_2+1.1605x_3-1.7695x_1^2+1.0589x_1x_2+2.1072x_2^2-0.9086x_2x_3+1.1985x_3^2\right\} \\ -1\le x_1x_2x_3\le 0.5 \end{cases}$$

and

$$\begin{cases} \max\left\{-23.6434+6.5660x_1+4.2935x_2-4.0217x_1x_2+2.8686x_1^2-2.3092x_2^2\right\} \\ -1\le x_1x_2\le 1 \end{cases}.$$

For the 2nd group:

$$\begin{cases} \min\left\{ 2.3146+1.0456x_1-1.2365x_2+0.8096x_3+0.6043x_1^2-0.5568x_1x_2+0.6152x_1x_3+0.3391x_2^2-0.6470x_2x_3-0.5324x_3^2\right\} \\ -1\le x_1x_2x_3\le 1 \end{cases}$$

and

$$\begin{cases} \max\left\{-23.0984+1.8192x_1+2.8881x_2-0.7176x_1^2-1.9597x_1x_2-0.6434x_1x_3-1.1643x_2^2-0.8429x_2x_3\right\} \\ -1\le x_1x_2x_3\le 1 \end{cases}.$$

Tables 1.7 and 1.8 give the optimal values for the standard deviation and the SNR of the 1st and the 2nd group. They also show the mean square error.

Table 1.7:

Optimal values for the standard deviation and the SNR for the 1st group.

	Optimum	Combinations			MSE
		x_1	x_2	x_3	
Stand. Dev.	0.60	0.5	0.5	-0.29	6.25
SNR	-14.20	1	0.06	$\forall x_3 \in R$	11.74

Table 1.8:

Optimal values for the standard deviation and the SNR for the 2nd group.

	Optimum	Combinations			*MSE*
		x_1	x_2	x_3	
Stand. Dev.	0.71	0.07	0.93	-1	4.25
SNR	-20.44	0.35	1	-1	4.83

The optimal values for the standard deviation in the first group and the second group are respectively 0.60 and 0.71. The optimal values for the SNR in both groups are respectively -14.20 and -20.44.

Optimal values for the mean and the variance

The optimization problem to be solved in order to obtain the optimal values for the mean and the variance is:

For the 1st group:

$$\begin{cases} \min\left\{\left(15.2917-11.7362x_1-11.9381x_2+6.6954x_2^2+10.9187x_1x_2\right)^2+10.89-9.29x_1-41.13x_2+16.68x_3-27.49x_1^2+ \right. \\ \qquad \left. +12.19x_1x_2-2.90x_1x_3+33.58x_2^2-28.33x_2x_3+23.55x_3^2\right\} \\ -1\leq x_1x_2x_3\leq 0 \end{cases}.$$

For the 2nd group:

$$\begin{cases} \min\left\{\left(14.2967-4.6106x_1-6.2331x_2+2.8535x_1^2+3.3335x_2^2+5.6862x_1x_2\right)^2+ \right. \\ \qquad \left. +9.12+7.56x_1-9.18x_2+4.30x_3+4.83x_1^2-8.63x_1x_2+4.04x_1x_3+3.52x_2^2-3.07x_2x_3-6.55x_3^2\right\}. \\ -1\leq x_1x_2x_3\leq 1 \end{cases}$$

Table 1.9 gives optimal settings, optimal values for the mean and variance models and the mean square error.

Table 1.9: Optimal values for the mean and variance.

Group	Optimal setting (x_1, x_2, x_3)	$\hat{y}_{mean}(x_{Opt})$	$\hat{y}_{Var}(x_{Opt})$	MSE
1	$(0, 0, -0.35)$	15.29	7.94	241.77
2	$(-0.75, -0.75, -1)$	29.11	0.05	847.34

Interpretation

Considering the analysis based on half fractional design , the pairs of optimal values for the standard deviation and the SNR corresponding to the first and the second groups are respectively $(0.60, -14.20)$ and $(0.71, -20.44)$. The corresponding pairs of optimal values for the mean and the variance are respectively $(15.29, 7.94)$ and $(29.11, 0.05)$.

The analysis based on half fractional design does not give the same information of the process. In fact, the number of impurities for the first group is 15.29%, and 29.11% for the second group. The discrepancy between both values is 13.82. The discrepancy between the optimal values for the variance in the first and the second groups is $7.94 - 0.05 = 7.89$.

1.4. Conclusions

In this Chapter, the theory of the Taguchi approach that relies on Double Orthogonal Array Design is shown. Related robustness metrics and an illustrative example are given. These metrics are the Quadratic Loss Function and the Signal-to-Noise Ratio. Some of the drawbacks of the Taguchi approach are summarized and references for further details are indicated. An overall application of the chemical process considering the full design analysis and fractional design analysis is given. The results are compared. It is verified that the process is improved by conducting an experimental design. In fact, the quality loss of not achieving the target is reduced while applying the optimal values obtained from the experiment.

Chapter 2: Combined Array Design

2.1. Overview of Combined Array Design

As mentioned in Chapter 1, one of the drawbacks of double orthogonal array design is that the number of runs required to conduct an experiment is generally large, and interactions between design and noise factors are not considered. Some authors go so far as saying that analysis related to double orthogonal array does not make proper use of the number of runs [22]. Combined array design is a single experimental design in control and noise factors. Both control and noise factors are then modelled, and the settings of the noise factors are no longer identical for each setting of the design factors. 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 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. Response surface methodology is used for determining optimal solutions for the mean and the variance.

2.2. Experimental structure

Suppose that the quality characteristic y of a product or process depends on p design factors $x_1,\dots,x_p$ and q noise factors $z_1,\dots,z_q$. The experimental structure of the combined array design is presented by Figure 2.1.

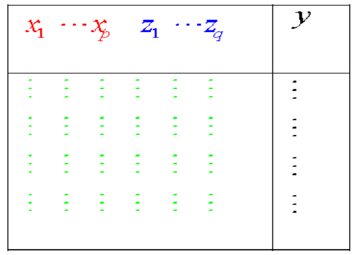


Figure 2.1: Experimental structure of Combined Array Design.

2.2.1. Dual Response Surface Approach

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

To explore the dependence of y on x and Z, the following model is assumed for the response, to accommodate control-by-noise interactions:

$$y(x,Z) = \beta_0 + \beta^T x + x^T B\, x + \gamma^T Z + x^T \Delta\, Z + \varepsilon \tag{2.1}$$

In this model, Z is the random noise vector, ε´s are independent identically distributed $N(0,\sigma^2)$ random errors. It is assumed that ε and Z are independent. The constant β_0, the vectors β, γ and the matrices B and Δ consist of unknown parameters, and σ^2 is also usually unknown. It is also assumed that $E(Z)=0$ and that $Cov(Z)=\Omega$ is known. The two response surfaces are obtained analytically from (2.1), one for the mean of y as a function of the control factors x, and one for the variance of y, also in terms of the control factors:

$$E_Z\left(y\left(x,Z\right)\right) = \beta_0 + \beta^T x + x^T B\, x\,. \tag{2.2a}$$

$$Var_Z\left(y\left(x,Z\right)\right) = \left(\gamma^T + x^T\Delta\right)\Omega\left(\gamma + \Delta\, x\right) + \sigma^2\,. \tag{2.2b}$$

In fact, $E_Z\left(y(x,Z)\right)=E_Z\left(\beta_0+\beta^T x+x^T B x+\gamma^T Z+x^T\Delta Z+\varepsilon\right)=\beta_0+\beta^T x+x^T B x$ and

$$
\begin{aligned}
Var_Z\left(y(x,Z)\right)&=Var_Z\left(\beta_0+\beta^T x+x^T B x+\gamma^T Z+x^T\Delta Z+\varepsilon\right)=Var_Z\left(\gamma^T Z+x^T\Delta Z+\varepsilon\right)=Var_Z\left(\gamma^T Z+x^T\Delta Z\right)+Var\left(\varepsilon\right)\\
&=Var_Z\left[\left(\gamma^T+x^T\Delta\right)Z\right]+\sigma^2=\left(\gamma^T+x^T\Delta\right)Var\left(Z\right)\left(\gamma^T+x^T\Delta\right)^T+\sigma^2=\left(\gamma^T+x^T\Delta\right)\Omega\left(\gamma+\Delta x\right)+\sigma^2.
\end{aligned}
$$

After model (2.1) is fitted to the data, $(2.2\,a)$ and $(2.2\,b)$ give two prediction models for the mean and variance of the process using model parameter estimates obtained from the fit, including estimating σ^2. In principle, model (2.1) could be extended to accommodate interactions among noise variables, quadratic terms in the noise variables and so on, but this requires the knowledge of higher moments of Z, and usually is not done [01].

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. Dominguez D. J and Ernesto Barrios Zamudio [09] have summarized and compared those methods. Myers R H and Carter W. H [24], and Myers R. H and Vining G. G. [25] 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 the 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:

$$
\begin{cases}
\text{optimize } \hat{E}_Z\left(y\left(x,Z\right)\right)\\
\hat{Var}_Z\left(y\left(x,Z\right)\right)=\sigma_0^2\\
x\in R
\end{cases}
. \qquad (2.2\,c)
$$

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

$$\begin{cases} \text{minimize } V\hat{a}r_Z\left(y\left(x,Z\right)\right) \\ \hat{E}_Z\left(y\left(x,Z\right)\right)=M \\ x\in R \end{cases} \qquad (2.2\,d)$$

R is the experimental region.

Dennis K. J. Lin and Wanzhu Tu [05] argue that this method of optimization 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 mean square error criterion. Their method will be used in Section.2.3.

2.3. Application of Dual Response Surface Approach

We adapt the application of the chemical process (John S. Lawson[19]) to the combined array design, specifically to the case of the dual response surface approach.

The data corresponding to the combined array design are given in Annex 2.1 and the results of analysis of variance are in Annex 2.2.

2.3.1. Statistical results

The idea is to fit the second order regression model
$y(x,Z)=\beta_0+\beta^T x+x^T Bx+\gamma^T Z+x^T\Delta Z+\varepsilon$ and then calculating
$\hat{E}_z\left(y\left(x,Z\right)\right)=\beta_0+\beta^T x+x^T Bx$ and $V\hat{a}r_z\left(y\left(x,Z\right)\right)=\left(\gamma^T+x^T\Delta\right)\Omega\left(\gamma+\Delta x\right)+\sigma^2$.
Annex 2.3 gives the coefficients of the regression response model.

Regression models

The results of analysis of variance for the response model are shown in Annex 2.2. The corresponding response model considering the significant effects is

$$y(x,Z) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{22} x_2^2 + \gamma_1 z_1 + \gamma_2 z_2 + \delta_{21} x_2 z_1 + \varepsilon .$$

The fitted response model is

$$\hat{y}(x,Z) = 14.7942 - 8.1734 x_1 - 9.0856 x_2 + 8.3025 x_1 x_2 + 5.0145 x_2^2 + 3.9057 z_1 - 1.2007 z_2 - 3.2969 x_2 z_1 .$$

The mean and variance models are:

$$\widehat{E}_z\left(y(x,Z)\right) = 14.7942 - 8.1734 x_1 - 9.0856 x_2 + 8.3025 x_1 x_2 + 5.0145 x_2^2 .$$

$$V\hat{a}r_z\left(y(x,Z)\right) = \left(2.7050 + 0.0543 x_1 - 2.2713 x_2\right)^2 + 13.54 = 20.857 + 0.294 x_1 - 12.288 x_2 - 0.247 x_1 x_2 + 0.003 x_1^2 + 5.159 x_2^2 .$$

Here, we replace $\widehat{\sigma^2}$ by the mean square error from the analysis of variance. We also assume that $z_i, i=1,2$ is a random variable with mean 0 and some variance σ_z^2 such that the levels of $z_i, i=1,2$ are at $\pm\sigma_z$ in coded form. Thus, $\sigma_z = 1$.

Optimal values

We apply the optimization method proposed by Dennis K. J. Lin and Wanzhu Tu $[05]$. The method consists of three steps:

Step 1: Find the fitted models $\widehat{E}_z\left(y(x,Z)\right)$ and $V\hat{a}r_z\left(y(x,Z)\right)$. Both are functions of x.

Step 2: Find x_{Opt} such that the mean square error given by the expression

$MSE = \left(\widehat{E}_z\left(y(x,Z)\right) - T\right)^2 + V\hat{a}r_z\left(y(x,Z)\right)$ is minimized. x_{Opt} is obtained by minimizing MSE subject to $x \in R$. In this application, $T = 0$.

Step 3: Evaluate $\widehat{E}_z\left(y(x_{Opt},Z)\right)$ and $V\hat{a}r_z\left(y(x_{Opt},Z)\right)$.

The optimization problem to be solved is:

$$\begin{cases} \min \quad \left\{\left(14.7942 - 8.1734 x_1 - 9.0856 x_2 + 8.3025 x_1 x_2 + 5.0145 x_2^2\right)^2 + \left(20.857 + 0.294 x_1 - 12.288 x_2 - 0.247 x_1 x_2 + 0.003 x_1^2 + 5.159 x_2^2\right)\right\} \\ -1 \le x_1 x_2 \le 1 \end{cases}.$$

Table 2.1 gives optimal settings, optimal values for the mean and variance models and the mean square error.

Table 2.1: Optimal values for the mean and variance.

Optimal setting (x_1, x_2, x_3)	$\widehat{E}_z\left(y(x_{Opt}, Z)\right)$	$V\hat{a}r_z\left(y(x_{Opt}, Z)\right)$	MSE
$(1, 0.16, \forall x_3)$	6.62	19.28	63.16

Interpretation

The analysis based on full design data shows that the number of impurities may be reduced to 6.62% with a variance of 19.28.

Application of the Quality Loss Function

We introduce in this chapter the illustration of the Quality Loss Function as shown in sub-section 1.2.5 of Chapter 1. We remember that the estimate of the expected loss is $\widehat{Q} = 25\left[\left(\widehat{\mu}\right)^2 + \widehat{\sigma^2}\right]$ By replacing $\widehat{\mu}$ and $\widehat{\sigma^2}$ respectively by $\widehat{E}_z\left(y(x_{Opt}, Z)\right)$ and $V\hat{a}r_z\left(y(x_{Opt}, Z)\right)$, the following quality loss is obtained:

$$\widehat{Q} = 25\left[\left(6.62\right)^2 + 19.28\right] = 1577.6 .$$

This value of the quality loss supports the idea given in Chapter 1 that the process may be improved by conducting a designed experiment and applying optimal values in the process. In fact, in this case, the reduction of the quality loss is $7476.25 - 1577.6 = 5898.6$ monetary units, for each unit of chemical product.

2.3.2. Analysis based on half fractional design

The two fractioned groups considered for this analysis are given in Annexes 2.4 and 2.7. The corresponding results for analysis of variance are in Annexes 2.5 and 2.8. The corresponding response models considering the significant effects are respectively:

$$y(x,Z)=\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+\beta_{11}x_1^2+\beta_{12}x_1x_2+\beta_{13}x_1x_3+\beta_{22}x_2^2+\beta_{23}x_2x_3+\beta_{33}x_3^2+\gamma_1z_1+\gamma_2z_2+\delta_{11}x_1z_1+\delta_{12}x_1z_2+$$
$$+\delta_{21}x_2z_1+\delta_{22}x_2z_2+\delta_{31}x_3z_1+\delta_{32}x_3z_2+\varepsilon$$

and

$$y(x,Z)=\beta_0+\beta_{11}x_1^2+\beta_{12}x_1x_2+\beta_{13}x_1x_3+\beta_{23}x_2x_3+\gamma_1z_1+\delta_{11}x_1z_1+\delta_{12}x_1z_2+\delta_{21}x_2z_1+\delta_{22}x_2z_2+\varepsilon\,.$$

The regression coefficients of the response models are in Annexes 2.3 and 2.6.
The fitted response models are respectively:

$$\hat{y}(x,Z)=15.4113+0.4585x_1-1.8341x_2+0.2228x_3+1.2470x_1^2+2.8418x_1x_2-3.4492x_1x_3+6.4518x_2^2-7.8108x_2x_3-3.2110x_3^2+$$
$$+9.9945z_1-0.7088z_2-5.2315x_1z_1+3.3225x_1z_2+1.1473x_2z_1+3.3320x_2z_2-1.0789x_3z_1-4.1690x_3z_2$$

and

$$\hat{y}(x,Z)=14.5195+5.7611x_1^2-8.5808x_1x_2+5.8355x_1x_3-7.1000x_2x_3-3.1343z_1-4.7043x_1z_1-3.9302x_1z_2-10.2783x_2z_1+$$
$$+6.1789x_2z_2.$$

The mean and variance models

For the 1st group:

$$\hat{E}_z\left(y(x,Z)\right)=15.4113+0.4585x_1-1.8341x_2+0.2228x_3+1.2470x_1^2+2.8418x_1x_2-3.4492x_1x_3+6.4518x_2^2-7.8108x_2x_3-3.2110x_3^2\,.$$

$$V\hat{a}r_z\left(y(x,Z)\right)=\left(9.2857-1.909x_1+4.793x_2-5.2479x_3\right)^2+115.393$$
$$=201.62-35.453x_1+89.013x_2-97.461x_3+3.6443x_1^2-18.300x_1x_2+20.036x_1x_3+22.93x_2^2-50.306x_2x_3+27.54x_3^2\,.$$

For the 2nd group:

$$\hat{E}_z\left(y(x,Z)\right)=14.5195+5.7611x_1^2-8.5808x_1x_2+5.8355x_1x_3-7.1000x_2x_3\,.$$

$$V\hat{a}r_z\left(y(x,Z)\right)=\left(-3.1343-8.6345x_1-4.0994x_2\right)^2+29.310$$
$$=39.134+54.126x_1+25.697x_2+74.555x_1^2+70.793x_1x_2+16.805x_2^2\,.$$

Optimal values

The optimization problem to be solved in order to obtain the optimal values for the mean and variance is:

For the 1st group:

$$\begin{cases} \min\left\{\left(15.4113+0.4585x_1-1.8341x_2+0.2228x_3+1.2470x_1^2+2.8418x_1x_2-3.4492x_1x_3+6.4518x_2^2-7.8108x_2x_3-3.2110x_3^2\right)^2+ \right. \\ \qquad \left. +201.62-35.453x_1+89.013x_2-97.461x_3+3.6443x_1^2-18.300x_1x_2+20.036x_1x_3+22.93x_2^2-50.306x_2x_3+27.54x_3^2\right\} \\ -1\le x_1x_2x_3\le 1 \end{cases}.$$

For the 2nd group:

$$\begin{cases} \min\left\{\left(14.5195+5.7611x_1^2-8.5808x_1x_2+5.8355x_1x_3-7.1000x_2x_3\right)^2+ \right. \\ \qquad \left. +39.134+54.126x_1+25.697x_2+74.555x_1^2+70.793x_1x_2+16.805x_2^2\right\}. \\ -1\le x_1x_2x_3\le 1 \end{cases}$$

Table 2.2 gives optimal setting, optimal values for the mean and variance models and the mean square error.

Table 2.2: Optimal values for the mean and the variance.

Group	Optimal setting (x_1,x_2,x_3)	$\hat{E}_z\left(y(x_{Opt},Z)\right)$	$V\hat{a}r_z\left(y(x_{Opt},Z)\right)$	MSE
1	$(1,0.37,1)$	9.05	17.24	212.44
2	$(-0.07,-1,-1)$	7.25	31.86	84.41

Interpretation

The pairs of optimal values of the number of impurities and the variance for the first and the second groups are respectively $(9.05,17.24)$ and $(7.25,31.86)$.

The results of the combined array design obtained in Chapter 2 are more homogeneous than those obtained in Chapter 1 where the analysis is based on double orthogonal array design. However, the conclusion given in the case of double orthogonal array design is still valid. This is, the analysis based on half fractional design does not give the same information of the process. In fact, considering the combined array design, the number of impurities obtained for the first group is 9.05%, and 7.25% for the second group. The discrepancy between both values is 1.8. This value seems to be small. However, the discrepancy between the optimal values for the variance in the first and the second groups is large: $31.86 - 17.24 = 14.62$.

In practice, it is recommended to conduct a confirmatory experiment in order to confirm the validity of the results previously obtained.

2.4. Modified Dual Response Surface Approach

2.4.1. Motivation

The existing procedures for robust design, devised for physical experiments, may be too limiting when the system can be simulated by a computer model. In this section we introduce a modification of the dual response surface modelling, which incorporates the option of stochastically simulating some of the noise factors when their probabilistic behaviour is known. The knowledge of the noise distribution may come either from historical data or from ad hoc measurements. In practice, internal noise factors are assumed to have a normal distribution. The method is also applicable in the case of crossed array design.

Alessandra Giovagnoli and Daniele Romano [01] stipulate that the method appears suitable for designing complex measurement system. They apply it to the design of a high- precision optical profilometer.

The purpose of this section is to illustrate theoretically, how stochastic simulation experiments can be best employed in the field of robust design.

Classical robust design relies on physical experiments whose factors are the parameters (control factors) and the noise (noise factors). Noise factors, even though they vary randomly in the process, they are controlled in the experiment. Designed experiments for robust design have been devised for physical settings, but in practical situations noise factors are typically difficult, if not impossible, to control. Thus, only a few factors, and with few levels, are usually included in the design.

This constraint can at times be relaxed in simulation. Simulated experiments have been performed by scientists and engineers, ever since the advent of the computer age, and are being increasingly used as an investigation tool in science and technology.

Within the framework of robust design, stochastic simulation appears to be a natural tool for transmitting distribution of noise input to the output. The advantage is to restore the centrality of randomness, which is the rationale for making an inference.

As already mentioned in this work, the statistical methodology underlying robust design, and now the most widely accepted, is the dual response surface methodology. Two surfaces are estimated, one for the mean and one for the variance of the process. The existing types of experiments for dual response models are the crossed array and the combined array. A new general protocol for conducting robust design studies on the computer extends the dual response surface approach. It is characterized by a different treatment of the noise factors, some of which are considered random, as they appear in the real process.

The method can be beneficial to the solution of an integrated parameter and tolerance design problem, by adding variances (or standard deviations) of internal noise factors as controls in the experiment and simulating the internal noise accordingly.

2.4.2. Theory of Modified Dual Response Surface Approach

Let us divide the random factors Z into two independent vectors, Z_1 and Z_2 where $Z_1 = (Z_1, \cdots, Z_m)$ and $Z_2 = (Z_{m+1}, \cdots, Z_q)$. The random vector Z_1 includes the variables that are simulated stochastically, whereas the remaining set of noise factors Z_2 are given fixed levels z_2 for some different choices of z_2. At the same time, different levels x of the control factors are also chosen for the experimentation. The computer experiment is performed by stochastically simulating the noise Z_1 for chosen pairs (x, z_2) and the sample mean and variance of the observed response are calculated.

The experimental structure of modified dual response surface is shown by Figure 2.2.

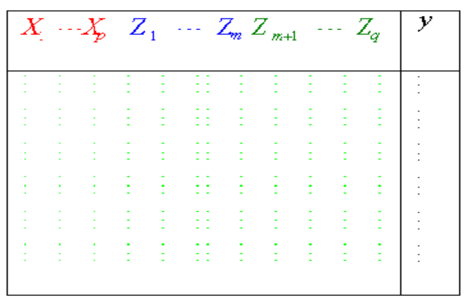


Figure 2.2: Experimental structure of Modified Dual Response Surface.

The following phase is to build a surrogate model of the simulator in the region of interest. The surrogate model is a probabilistic model that describes the experiment, similarly to what happens for physical experiments. To start with, we assume additivity between a systematic effect y_1 on response y purely due to the control factor x and a random effect Y_2 due to Z and its interactions with x: $y = y_1(x) + y_2$.

Furthermore, it is assumed that $y_1(x)$ is linear in unknown parameters β: $y_1(x) = f(x)^T \beta$ where $f(x)$ is a known vector function.

As to the random term y_2, it is assumed that, when Z_2 is fixed, its expectation over the noise vector Z_1 is also linear in another set of parameters γ.

Thus,

$$E_{Z_1}\left(y \mid Z_2 = z_2\right) = f(x)^T \beta + g\left(x, z_2\right)^T \gamma \ ; \qquad (2.3\,a)$$

where $g\left(x, z_2\right)$ is again a known vector function.

Assume a linear model in a set of parameters δ for the log-variance of y, again when Z_2 is fixed and the variance is taken over Z_1.

Then,

$$\log Var_{Z_1}\left(y \mid Z_2 = z_2\right) = h\left(x, z_2\right)^T \delta \ ; \qquad (2.3\,b)$$

with $h\left(x, z_2\right)$ a known vector function.

Stage 1: Estimating the unknown parameters β, γ, δ

For every $\left(x, z_2\right)$ in the experiment, $E_{Z_1}\left(y \mid Z_2 = z_2\right)$ and $Var_{Z_1}\left(y \mid Z_2 = z_2\right)$ are observed via the corresponding sample mean and sample variance obtained under the simulation. Then, the parameters β and γ in $(2.3\,a)$ are estimated by weighted least squares, since variances at different experimental points are different, and δ in $(2.3\,b)$ is estimated by ordinary least squares.

Stage 2: Building predictors

The fitted dual model $(2.3\,a)+(2.3\,b)$ describes the mean and the variance of the response for the given levels of the noise factors Z_2, controlled in the experiment but random in the process. Thus, this fitted model, in general, is not very interesting in itself. However, the in-process mean and variance are related to $(2.3\,a)$ and $(2.3\,b)$ by means of the identities:

$$E_Z(y) = E_{Z_2}\left[E_{Z_1}(y \mid Z_2)\right]. \qquad (2.4\,a)$$

$$Var_Z(y) = E_{Z_2}\left[Var_{Z_1}(y \mid Z_2)\right] + Var_{Z_2}\left[E_{Z_1}(y \mid Z_2)\right]. \qquad (2.4\,b)$$

Equation $(2.4\,a)$ yields

$$E_Z(y) = f(x)^T \beta + E_{Z_2}\left[g(x, z_2)\right]^T \gamma = f(x)^T \beta + a(x)^T \gamma; \qquad (2.5\,a)$$

where $a(x)$ is a vector function that can either be computed analytically, since $g(x, z_2)$ and the probability distribution of the noise vector Z_2 are assumed known or can be computed numerically. The in-process mean $E_Z(y)$ is estimated substituting the estimates of β and γ in expression $(2.5\,a)$.

Hence, Equation $(2.4\,b)$ becomes

$$Var_Z(y) = E_{Z_2}\left[\exp\left\{h(x, Z_2)^T \delta\right\}\right] + Var_{Z_2}\left[g(x, Z_2)^T \gamma\right] = b(x, \delta) + \gamma^T C(x)\gamma; \qquad (2.5\,b)$$

where the function $b(x, \delta) = E_{Z_2}\left[\exp\left\{h(x, Z_2)^T \delta\right\}\right]$ and $C(x) = Cov_{Z_2} g(x, Z_2)$ are computed similarly to $a(x)$. The in-process variance of y is estimated substituting the estimates of γ and δ in expression $(2.5\,b)$.

Compared with the crossed array, the modified dual response surface approach provides additional information on how noise in Z_2 affects process variance, via the functions g and h.

The modified dual response surface may be generalized in a multi-step approach if the noise factors are simulated by sequentially adding one of them at each step in the Z_1 vector. This allows one to sequentially evaluate how an individual noise factor affects overall variability.

Special cases

The special cases of the modified dual response surface are particular behaviours of noise factors that lead to double orthogonal array design or to dual response surface.

1. When $E_{Z_2}\left[g\left(x,Z_2\right)\right]=0$, the in-process mean $(2.5\,a)$ becomes $E_Z\left(y\right)=f(x)^T\beta$ which is already estimated at Stage 1. This occurs in the combined array model.

In fact, $E_Z\left(y\right)=f(x)^T\beta+E_{Z_2}\left[g\left(x,Z_2\right)\right]^T\gamma=f(x)^T\beta$ because $E_{Z_2}\left[g\left(x,Z_2\right)\right]^T=0$.

2. In the case $Z_1=\varnothing$ (or equivalently, $Z_2=Z$), the simulation becomes non-stochastic. Expressions $(2.3\,a)$ and $(2.3\,b)$ collapse to just one equation: $y=f(x)^T\beta+g(x,Z)^T\gamma$. In particular, if the vector $f(x)^T\beta$ is a second degree polynomial in the x's and $g(x,Z)^T\gamma$ contains only linear terms in the z's and xz-product terms, model (2.3) reduces to model (2.1) for the combined array approach, but without the experiment error.

In fact,

- $E_{Z_1}\left(y\,|\,Z_2=z_2\right)=E\left(y\,|\,Z\right)=f(x)^T\beta+g\left(x,Z\right)^T\gamma$.
- There is no need of calculating $\log Var_{Z_1}\left(y\,|\,Z_2=z_2\right)$ as $Z_1=\varnothing$.
- Models (2.1) and (2.3) are:

$$y(x,Z)=\beta_0+\beta^T x+x^T B\,x+\gamma^T Z+x^T\Delta Z+\varepsilon.$$

$$E_{Z_1}\left(y\,|\,Z_2=z_2\right)=f(x)^T\beta+g\left(x,z_2\right)^T\gamma.$$

Then $f(x)^T\beta\cong\beta_0+\beta^T x+x^T B\,x$ and $g(x,Z)^T\gamma\cong\gamma^T Z+x^T\Delta Z$.

3. In the case $Z_1=Z$ (or equivalently, $Z_2=\varnothing$) $(2.3\,a)$ and $(2.5\,a)$ imply

$E_{Z_1}\left(y\right)=a(x)^T\gamma$.

Expressions $(2.3\,a)$ and $(2.3\,b)$ become

$$E_Z\left(y\right)= f(x)^T\beta + a(x)^T\gamma \qquad (2.6\,a)$$

together with

$$\log Var_Z\,Y\left(x,Z\right)=h(x)^T\delta. \qquad (2.6\,b)$$

When the noises Z are independent and normally distributed, one obtains the same dual response surface model from $\left(2.6\,a\right)$ and $\left(2.6\,b\right)$ as from $y_{ij} = f\left(x_i\right)^T\beta+\varepsilon_{ij}\,\sigma_i$ and $\log s_i^2 = h\left(x_i\right)^T\delta+\varepsilon_i^*$ in the crossed array approach $\left[01\right]$.

In fact,

- $$\begin{cases} E_{Z_1}\left(y\,|\,Z_2=z_2\right)= f(x)^T\beta+g\left(x,z_2\right)^T\gamma & (2.3\,a)\\ E_Z\left(y\right)= f(x)^T\beta+E_{Z_2}\left[g\left(x,Z_2\right)\right]^T\gamma= f(x)^T\beta+a(x)^T\gamma & (2.5\,a)\end{cases}$$

Equations $\left(2.3\,a\right)$ and $\left(2.5\,a\right)$ imply $E_{Z_1}\left(y_2\right) \underset{Z_1=Z}{=} E_Z\left(y_2\right)=a(x)^T\gamma$.

Equations $\left(2.3\,a\right)$and $\left(2.5\,a\right)$ imply again $E_{Z_1}\left(y\right) \underset{Z_1=Z}{=} E_Z\left(y\right)= f(x)^T\beta+a(x)^T\gamma$ and

$\log Var_{Z_1}\left(y\,|\,Z_2=z_2\right)=h\left(x,z_2\right)^T\delta$ leads to $\log Var_Z\left(y\right)=h\left(x\right)^T\delta$ from the fact that $Z_1=Z$.

2.5. Conclusions

In this Chapter, the theory of combined array design is exposed. An illustrative example of dual response surface approach is given. The results are compared to those obtained from the double orthogonal array design in Chapter 1. It seems realistic to analyse the data of the chemical process for both complete design and fractional design by the combined array design structure. In fact, the discrepancy between optimal values is less pronounced in the case of combined array design. This result supports the idea invoked by Kunert J. et al $\left[22\right]$ that the product array design does not make proper use of the number of runs. Through an illustrative calculus, we verify the assertion that conducting an experiment and applying the combination of the levels of the design factors that give the optimal values reduce the cost of the process.

Chapter 3:

Integrated method of Parameter design and Tolerance design

3.1. Generalities

Sometimes robust design aims at setting both design parameters and tolerance specifications. While specifying tolerances on a design parameter, the designer is actually putting a limit on the random variability of the parameter, i.e. on the standard deviation of the corresponding internal noise factor [01]. A tighter tolerance around the nominal value results in a smaller transmitted variation on the response, but incurs an extra cost [34]. The integrated method of parameter design and tolerance design aims at choosing nominal values of design factors and their tolerances simultaneously so that the cost of the process is minimized.

3.2. Parameter design and Tolerance design

Consider a process where the response of interest y depends on the factors $x^T = \left(x_1, x_2, \cdots, x_p\right)$ expressed in terms of original levels. The corresponding factors in coded levels are $w^T = \left(w_1, w_2, \cdots, w_p\right)$.

The second order response surface model in terms of coded factors is:

$$y\left(w\right) = \alpha_0 + w^T\alpha + w^T \mathrm{A}\, w + \varepsilon . \tag{3.1}$$

Coded factors and original factors are linked by the expression

$$w_i = \frac{x_i - \bar{x}_i}{\frac{1}{2}R_{x_i}}, \; i = 1, 2, \cdots, p . \tag{3.2}$$

The model in terms of original factors becomes:

$$y(x)=\beta_0+x^T\beta+x^T\Delta x+\varepsilon \tag{3.3}$$

where β_0 is a constant, $\beta^T=(\beta_1,\beta_2,\cdots,\beta_p)$ is a vector of unknown parameters, $\Delta=[\delta_{ij}]$ is a $p\times p-$matrix of the second order parameters, (that is $\delta_{ij}=\beta_{ii}$ if $i=j$ and $\delta_{ij}=\frac{1}{2}\beta_{ij}$ if $i\neq j$), $\varepsilon\sim N(0,\sigma_\varepsilon^2)$.

The levels of the factors are fixed in the experimental phase. However, during the production process, it is common that, instead of observing x_i, one observes x_i+t_i. The term t_i is a random variable referred to as tolerance, associated with the factor x_i such that $E(t_i)=0$, $Var(t_i)=\sigma_i^2$ and $\operatorname{cov}(t_i,t_j)=\rho_{ij}\sigma_i\sigma_j$.

The model that includes the factors and their corresponding tolerances is:

$$y(x\,|\,t)=\beta_0+(x+t)^T\beta+(x+t)^T\Delta(x+t)+\varepsilon. \tag{3.4}$$

Suppose that the experimenter wants the response of interest y to reach the target M .

The quality loss function that indicates the cost caused by not achieving the

target is:

$$L(y(x))=k(y(x)-M)^2; \tag{3.5}$$

where the coefficient k is a quality loss constant associated with each unit product. The average quality loss is:

$$Q=E[L(y(x))]=E\left[k(y(x)-M)^2\right]=k\left\{Var(y(x))+[E(y(x))-M]^2\right\}. \tag{3.6}$$

Let $C_i(t_i)$ represent the cost function due to applying the tolerance t_i to the factor x_i. Then the total cost incurred by all of the tolerances is:

$$C(t)=\sum_{i=1}^{p}C_i(t_i). \tag{3.7a}$$

The cost functions are determined by the collected information of the manufacturing cost versus the tolerance for each of the control factors. It is obvious that a higher precision level with tighter tolerance usually requires a higher manufacturing cost. There is a monotonic decreasing relationship between manufacturing cost and tolerance. In literature, several models are reported to describe the cost-tolerance relationship. Table 3.1 summarizes some of the cost functions as they are described by Joe Meng et al. [18].

Table 3.1: Cost-tolerance functions.

Cost-tolerance function	Form
The Sutherland function	$C(t) = \beta t^{-\alpha}$
Reciprocal square function	$C(t) = \frac{\alpha}{t^2}$
Reciprocal function	$C(t) = \frac{\alpha}{t}$
Exponential function	$C(t) = \alpha \exp(-bt)$
Michael-Siddall function	$C(t) = \alpha t^{-\beta} \exp(-\gamma t)$

Domínguez D. J. [08] proposes the following reciprocal cost function:

$$C_i(t_i) = \alpha + \beta t_i^{-\gamma}. \quad (3.7\text{ b})$$

This function will be used in Section 3.3.

The coefficients α, β and γ are determined by applying non linear regression methods.

Assume that each x_i follows Normal distribution $N\left(x_{i_0},\sigma_i^2\right)$ where x_{i_0} is the nominal value of the i th input variable x_i. As mentioned by Domínguez D. J. $[08]$ and William Li and C. F. J. Wu $[34]$, the tolerance t_i of x_i is:

$$t_i = 3\sigma_i \ . \tag{3.8}$$

Let $x_0 = \left(x_{1_0}, x_{2_0}, \cdots, x_{p_0}\right)$ and $t = \left(t_1, t_2, \cdots, t_p\right)$ be respectively the vectors of nominal values and tolerances. Because the normally distributed x_i is determined by its mean x_{i_0} and its standard deviation σ_i, the average quality loss defined by (3.6) is a function of x_{i_0} and σ_i, or equivalently of x_{i_0} and t_i. The average quality loss

may be written as:

$$Q = Q\left(x_0, t\right). \tag{3.9}$$

The total cost becomes:

$$H\left(x_0,t\right) = Q\left(x_0,t\right) + C(t) = k\left\{Var\left(y\left(x\,|\,t\right)\right) + \left[E\left(y\left(x\,|\,t\right)\right) - M\right]^2\right\} + \sum_{i=1}^{p} C_i\left(t_i\right). \tag{3.10}$$

Consider the model $y\left(x\,|\,t\right) = \beta_0 + \left(x+t\right)^T \beta + \left(x+t\right)^T \Delta\left(x+t\right) + \varepsilon$.

Assuming the tolerances are uncorrelated, the mean of $y\left(x\,|\,t\right)$ is:

$$E\left(y\left(x\,|\,t\right)\right) = \beta_0 + x^T\beta + x^T\Delta\, x + tr\Delta\Sigma\,; \tag{3.11 a}$$

where $\Sigma = \left[\sigma_i^2\right], i = 1, 2, \cdots, p$ is a diagonal matrix.

Then,

$$tr\Delta\Sigma = \sum_{i=1}^{p} \beta_{ii}\sigma_i^2 = \sum_{i=1}^{p} \beta_{ii}\left(\frac{t_i}{3}\right)^2 = \frac{1}{9}\sum_{i=1}^{p} \beta_{ii} t_i^2 \ . \tag{3.11 b}$$

Finally,

$$E\left(y\left(x\,|\,t\right)\right) = E\left(y\left(x\right)\right) + \frac{1}{9}\sum_{i=1}^{p}\beta_{ii}t_i^2 = \beta_0 + x^T\beta + x^T B\, x + \frac{1}{9}\sum_{i=1}^{p}\beta_{ii}t_i^2 \ . \tag{3.11 c}$$

In fact, $y(x|t)=\beta_0+(x+t)^T\beta+(x+t)^T W(x+t)+\varepsilon=\beta_0+x^T\beta+t^T\beta+x^T\Delta x+x^T\Delta t+t^T\Delta x+t^T\Delta t+\varepsilon$

Then, $E(y(x|t))=E\left(\beta_0+x^T\beta+t^T\beta+x^T\Delta x+x^T\Delta t+t^T\Delta x+t^T\Delta t\right)=\beta_0+x^T\beta+x^T\Delta\, x+E\left(t^T\Delta t\right)$. The expression $E\left(t^T\Delta t\right)$ leads to $E\left(t^T\Delta t\right)=tr(\Delta\Sigma)+(E(t))^T\,\Delta\,E(t)=\sum_{i=1}^{p}\beta_{ii}\sigma_i^2$ where

$$\Delta\Sigma=\begin{bmatrix}\beta_{11} & \frac{1}{2}\beta_{12} & \cdots & \frac{1}{2}\beta_{1p}\\ \vdots & \beta_{22} & & \vdots\\ & & \ddots & \\ \frac{1}{2}\beta_{p1} & & \cdots & \beta_{pp}\end{bmatrix}\begin{bmatrix}\sigma_1^2 & 0 & \cdots & 0\\ \vdots & \sigma_2^2 & & \vdots\\ & & \ddots & \\ 0 & & \cdots & \sigma_p^2\end{bmatrix}=\begin{bmatrix}\beta_{11}\sigma_1^2 & \frac{1}{2}\beta_{12}\sigma_2^2 & \cdots & \frac{1}{2}\beta_{1p}\sigma_p^2\\ \vdots & \beta_{22}\sigma_2^2 & & \vdots\\ & & \ddots & \\ \frac{1}{2}\beta_{p1}\sigma_1^2 & \cdots & & \beta_{pp}\sigma_p^2\end{bmatrix}. \quad (3.11\text{ d})$$

The variance of $y(x|t)$ is:

$$Var\left(y(x|t)\right)=(\beta+2\Delta\,x)^T\,\Sigma\,(\beta+2\Delta\,x)+\sigma_\varepsilon^2. \quad (3.12\text{ a})$$

In terms of tolerances, the variance becomes:

$$Var\left(y(x|t)\right)=\frac{1}{9}(\beta+2\Delta\,x)^T\,T\,(\beta+2\Delta\,x)+\sigma_\varepsilon^2\ . \quad (3.12\text{ b})$$

We remember that $\Delta=\left[\delta_{ij}\right]$ where $\delta_{ij}=\beta_{ii}$ if $i=j$ and $\delta_{ij}=\frac{1}{2}\beta_{ij}$ if $i\neq j$, $T=\left[t_i^2\right]$ is a diagonal matrix of tolerances. The final expression of the total cost $H(x_0,t)=Q(x_0,t)+C(t)$ becomes:

$$H(x_0,t)=k\left\{\frac{1}{9}(\beta+2\Delta\,x)^T\,T\,(\beta+2\Delta\,x)+\sigma_\varepsilon^2+\left[\beta_0+x^T+x^T\Delta x+\frac{1}{9}\sum_{i=1}^{p}\beta_{ii}t_i^2-M\right]\right\}^2+\sum_{i=1}^{p}C_i(t_i). \quad (3.13)$$

The optimization problem to be solved in order to obtain the setting that minimizes the total cost of the process is:

$$\begin{cases}\min H(x_0,t)\\ \text{subj. to } x_0\\ \qquad\qquad t\end{cases}. \quad (3.14)$$

3.3. Application of Parameter design and Tolerance design

This application is adopted from Raymond H. Myers and Douglas C. Montgomery [28]. One step in the production of a particular polyamide resin is the addition of amines. It is supposed that the manner of addition has a profound effect on the molecular weight distribution of the resin. Three variables are thought to play a major role: temperature at the time of addition (x_1, °C), agitation (x_2, RPM), and rate of addition (x_3, 1/min). Because it is difficult to physically set the levels of addition and agitation, three levels are chosen and a Box-Behnken Design is used. The viscosity of the resin is recorded as an indirect measure of molecular weight. The data, including the factors and their levels, the design, and the response values are given in Annex 3.1. The results of analysis of variance for the viscosity of the resin are given in Annex 3.2.

3.3.1. Results

The estimated regression model is:

$$\hat{y}(x) = -58.875 + 2.65x_1 - 0.65x_2 - 11.125x_3 - 0.012x_1^2 + 0.3x_2^2 - 0.145x_3^2 - 0.032x_1x_2 + 0.088x_1x_3 + 0.14x_2x_3 .$$

Let $t_i \sim N\left(0, \sigma_i^2\right), i = 1, 2, 3$.

From the estimated regression model, the matrix Δ is:

$$\Delta = \begin{pmatrix} \beta_{11} & \frac{1}{2}\beta_{12} & \frac{1}{2}\beta_{13} \\ \frac{1}{2}\beta_{21} & \beta_{22} & \frac{1}{2}\beta_{23} \\ \frac{1}{2}\beta_{31} & \frac{1}{2}\beta_{32} & \beta_{33} \end{pmatrix} = \begin{pmatrix} -0.012 & -0.016 & 0.044 \\ -0.016 & 0.3 & 0.07 \\ 0.044 & 0.07 & -0.145 \end{pmatrix} .$$

The matrix Σ is defined as $\Sigma = \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{pmatrix}$.

The following phase is to calculate $\hat{E}\left(y(x|t)\right)$ and $V\hat{a}r(y(x|t)$.

We apply respectively (3.11 a) and (3.12 a).

After simplification, the following result for $\widehat{E}\left(y(x\,|\,t)\right)$ is obtained:

$$\widehat{E}\left(y(x\,|\,t)\right)=-58.875+2.65x_1-0.65x_2-11.125x_3-0.012x_1^2+0.3x_2^2-0.145x_3^2-0.032x_1x_2+0.088x_1x_3+ \\ +0.14x_2x_3-0.012\sigma_1^2+0.3\sigma_2^2-0.145\sigma_3^2\ .$$

By taking into account the expression (3.8) which implies that $\sigma_i=\frac{t_i}{3}$, $\widehat{E}\left(y(x\,|\,t)\right)$ is transformed into:

$$\widehat{E}\left(y(x\,|\,t)\right)=-58.875+2.65x_1-0.65x_2-11.125x_3-0.012x_1^2+0.3x_2^2-0.145x_3^2-0.032x_1x_2+0.088x_1x_3+ \\ +0.14x_2x_3-0.0013t_1^2+0.0333t_2^2-0.0161\,t_3^2\ .$$

The estimate of $Var(y(x\,|\,t)=\left(\beta+2\Delta x\right)^T\Sigma\left(\beta+2\Delta x\right)+\sigma_\varepsilon^2$ is:

$$V\hat{a}r(y(x\,|\,t)=\left(2.65-0.024x_1-0.032x_2+0.088x_3\right)^2\sigma_1^2+\left(-0.65-0.032x_1+0.6x_2+0.14x_3\right)^2\sigma_2^2+ \\ +\left(-11.25+0.088x_1+0.14x_2-0.29x_3\right)^2\sigma_3^2+\widehat{\sigma}_\varepsilon^2\ .$$

Let us consider the following values for the parameters that appear in $\widehat{E}(y(x\,|\,t)$ and $V\hat{a}r(y(x\,|\,t)$: $\sigma_1=3, \sigma_2=0.15, \sigma_3=0.5, \widehat{\sigma}_\varepsilon^2=10.25$ which imply that $t_1=9, t_2=0.45, t_3=1.5$.

Let the target be $M=55$ and the quality loss for unit $k=1$.

The tolerances of the control factors and the related costs are given in Annex 3.3. The cost function (3.7 b) is used and the adjusted cost functions are:

$C(t_1)=0.132+1.9474\times t_1^{-0.6051}, C(t_2)=0.141+0.3956\times t_2^{-0.7820}$ and $C(t_3)=0.106+0.5280\times t_3^{-0.8173}$.

In order to formulate the optimization problem, the following expressions are calculated:

- $\widehat{E}\left(y(x\,|\,t)\right)=-59.013+2.65x_1-0.65x_2-11.125x_3-0.012x_1^2+0.3x_2^2-0.145x_3^2-0.032x_1x_2+0.088x_1x_3+0.14x_2x_3$
- $\left(\widehat{E}\left(y(x\,|\,t)\right)-55\right)^2=\left(-114.013+2.65x_1-0.65x_2-11.125x_3-0.012x_1^2+0.3x_2^2-0.145x_3^2-0.032x_1x_2+0.088x_1x_3+0.14x_2x_3\right)^2$
- $V\hat{a}r\left(y(x\,|\,t)\right)=105.1-11.942x_1-2.331x_2+5.825x_3+0.520x_1^2+0.022x_2^2+0.091x_3^2+0.143x_1x_2-0.393x_1x_3-6.7208x_2x_3$
- $C(t)=\sum_{i=1}^{3}C(t_i)=0.132+1.9474\left(9\right)^{-0.6051}+0.141+0.3956\left(0.45\right)^{-0.7820}+0.106+0.5280\left(1.5\right)^{-0.8173}=2.012\ .$

The optimization problem to be solved is:

$$\begin{cases} \min \ H(x,t) \\ x \in R \end{cases},$$

where $H(x,t) = \left(\widehat{E}\left(y(x\,|\,t)\right) - 55\right)^2 + V\hat{a}r\left(y(x\,|\,t)\right) + C(t)$, $150 \le x_1 \le 200$, $5 \le x_2 \le 10$ and $15 \le x_3 \le 25$.

The solution of the problem shows that the total cost of the process is 7475.698 monetary units.

The corresponding combination of the levels of the design factors is

$(x_1, x_2, x_3) = (150, 10, 25)$.

After the combination of the levels of the design factors is obtained; the expected mean, the variance and the expected value of the quality loss of the process are calculated using respectively the formulas (3.11 c), (3.12 a) and (3.6).

Then $\widehat{E}\left(y(x\,|\,t)\right) = 40.2370, V\hat{a}r\left(y(x\,|\,t)\right) = 7\,255.7$ and $Q = 7\,473.7$.

Interpretation

By optimizing the total cost incurred by parameter design and tolerances of the design factors, the combination of the levels of the factors that lead to the minimal cost is $(x_1, x_2, x_3) = (150, 10, 25)$.

The corresponding values of the expected mean and variance are respectively 40.2370 and 7255.7. The quality loss of the process is 7473.7 monetary units.

3.4. Conclusions

In this Chapter, the integrated method of parameter design and tolerance design is exposed. The method consists of integrating the tolerances related to nominal values of the levels of the design factors, and then minimizing the total cost of the process. The total cost of the process is composed of the quality loss and the cost of incorporating tolerances in the process. Even though the method causes extra cost in the process, it is the best way of ensuring the minimal variability once the nominal values of the design factors are obtained.

Chapter 4: Split-Plot Designs for Robust Experimentation

4.1. General view of Split-Plot Designs

In some multifactor factorial experiments it is impossible to completely randomize the order of the runs. Industrial experiments often encounter such situations where some experimental factors are hard to change or where there is significant discrepancy in the size of some experimental units [29]. In this case, the experimenters fix the levels of the difficult-to-change factors, and run all the combinations of the other factors. This leads to a split-plot design.

Soren Bisgaard [30] reveals another situation where split-plot designs are applied. This is the situation where all factors are equally difficult or easy to change and there are no other practical or economical restrictions on the randomization.

Split-plot designs have their origins in agricultural experiments. In these experiments, a factor such as the irrigation method is randomly applied to large sections of land, called whole plots. These sections then are split into smaller plots called subplots, and another factor, such as different fertilizers, is applied in random order within the subplots.

In the case of robust design, the distinction between design and noise factors may not be the reason for imposing a restriction on the randomization and using split-plot designs. The random character of the noise factors permits conducting robust design in split-plot design [30].

The analysis of a split-plot design consists of two error terms because there are two types of experimental units. These are the experimental units for the whole-plot factors and the experimental units for the sub-plot factors.

George E. P. Box and Stephen Jones [14] provide an extensive overview of split-plot and strip-plot designs for robust product experimentation. They consider three types of split-plot designs for robust experimentation with regard to the arrangement of the design and noise factors. Soren Bisgaard [30] provides specific technical details on how to design and analyse split-plot designs when they are based on combinations of two-level factorials and fractional factorials. His focus is on how to design and analyse $2^{k-p} \times 2^{q-r}$ split-plot designs. Soren also proposes the division of the sources of variation into two parts, those coming from the whole-plots and those coming from the sub-plots, and then constructs a normal plot for each group. This separation permits observing the factors and the interactions between factors which have effects on the response of interest

4.2. Arrangement of the design factors and the noise factors

In the case of split-plot designs for robust experimentation, three experimental arrangements of the design and the noise factors are of the most interest [14].

Arrangement 1: Design factors are split-plot factors.

In this arrangement, the whole-plots contain the noise factors and the sub-plots contain the design factors.

Arrangement 2: Noise factors are split-plot factors.

The whole-plots contain the product design factors and the sub-plots contain the noise factors.

Arrangement 3: Strip-block designs.
In this arrangement, the subplot treatments are assigned randomly in strips across each block of whole-plot treatments. In agricultural designs, this arrangement is frequently referred to as a strip-block experiment.

In this chapter, we pay more attention to the case where the noise factors are the whole-plot factors. The design allows fitting a first-order model with interactions. The general structure of the model is:

$$y = \text{WP factors} + \text{WP}_{\text{Error}} + \text{SP factors} + \text{WP}\times\text{SP interactions} + \text{SP}_{\text{Error}} \tag{4.1}$$

where y is the response variable, WP represents whole-plot and SP represents sub-plot.

4.3. Application

We adapt the problem of chemical process (John S. Lawson [19]). The aim of this application is to analyse the data as split-plot design. The results are compared to those obtained in the case of double orthogonal array design and combined array design.

We consider the arrangement where the whole-plots contain the noise factors and the sub-plots contain the design factors. This case seems to be more realistic because of the random nature of the noise factors.

There are 2 noise factors z_1 and z_2 applied to the whole-plots and 3 design factors x_1, x_2 and x_3 applied to the sub-plots. There is no replication.

In order to adapt the original design to the split-plot design, we combine in pairs the levels of the noise factors to form a whole-plot. There are four whole-plots. In practice, once the whole plot is randomly selected, the following step is to apply the combinations of the design factors within the whole-plot.

Table 4.1 shows the arrangement structure of the data in split-plot design. Annexes 4.1, 4.2 and 4.4 contain the data respectively of the full split-plot design and half fractional split-plot design for the first and the second groups.

The first-order model including two-factor interactions for the split-plot design is:

$$\begin{aligned} y(x,Z) = {} & \beta_0 + \beta_1 Z + WP_{Error} + \gamma_1 x_1 + \gamma_2 x_2 + \gamma_3 x_3 + \gamma_{12} x_1 x_2 + \gamma_{13} x_1 x_3 + \gamma_{23} x_2 x_3 + \\ & + \beta_{11} Z\, x_1 + \beta_{12} Z\, x_2 + \beta_{13} Z\, x_3 + SP_{Error} \ . \end{aligned} \tag{4.2}$$

Table 4.1: Arrangement of the data in split-plot design structure.

Noise factors	Levels			
z_1	1	-1	-1	1
z_1	-1	-1	1	1
Z	-1	1	-1	1
Whole-plot	1	2	3	4

Table 4.2 gives the coefficients of the regression models.

Table 4.2: Coefficients of the regression models.

	Intercept	Linear			Interaction					
Design	β_0	γ_1	γ_2	γ_3	γ_{12}	γ_{13}	γ_{23}	β_{11}	β_{12}	β_{13}
Full design	17.84	-8.17	-9.08	-0.13	8.30	0.07	0.17	-0.08	-1.68	0.76
1st group	18.75	-8.83	-10.49	0.36	10.92	-0.81	-1.41	0.61	-1.83	0.85
2nd group	17.01	-8.83	-8.34	-0.63	5.69	0.96	1.76	-0.76	-1.53	0.67

Table 4.3 shows the results of the analysis of variance for the data in complete split-plot design. The results of the analysis of variance corresponding to the first group and the second group of half fractional split-plot design are indicated by Annexes 4.3 and 4.5.

Table 4.3:

Analysis of variance for the number of impurities considering full design.

Source	DF	SS	Adj.SS	Adj. MS	F	P
Z	1	275.89	275.89	275.89	0.55	0.535
WP(Z)	2	1001.75	1001.75	500.87	23.84	0.000
x1	1	2137.76	2137.76	2137.76	101.74	0.000
x2	1	2641.55	2641.55	2641.55	125.71	0.000
x3	1	0.58	0.58	0.58	0.03	0.869
x1*x2	1	1102.90	1102.90	1102.90	52.49	0.000
x1*x3	1	0.09	0.09	0.09	0.00	0.949
x2*x3	1	0.50	0.50	0.50	0.02	0.878
Z*x1	1	0.21	0.21	0.21	0.01	0.920
Z*x2	1	90.38	90.38	90.38	4.30	0.044
Z*x3	1	18.47	18.47	18.47	0.88	0.353
Error	47	987.58	987.58	21.01		
Total	59	8257.68				

Stand. error = 4.58393
R-squar. = 88.04%
R-squar.(adjust.) = 84.99%

Table 4.4 presents the fitted regression models in terms of significant effects.

Table 4.4: Regression models.

Design	Regression model
Full design	$\hat{y}(x,Z)=17.84-8.17x_1-9.08x_2+8.30x_1x_2+0.76Z\,x_3$
1st group	$\hat{y}(x,Z)=18.75-8.83x_1-10.49x_2+10.92x_1x_2$
2nd group	$\hat{y}(x,Z)=17.01-8.83x_1-8.34x_2+5.69x_1x_2$

Optimal values for the number of impurities

Let us assign the low level to the whole-plot in the regression model related to the full design, this is $Z = -1$.
The optimization problem to be solved in order to obtain a minimal value for the number of impurities is:

For the full design:

$$\begin{cases} \min\left\{ 17.84 - 8.17x_1 - 9.08x_2 - 0.76x_3 + 8.30x_1x_2 \right\} \\ -1 \le x_1x_2x_3 \le 1 \end{cases}.$$

For the 1st group:

$$\begin{cases} \min\left\{18.75 - 8.83x_1 - 10.49x_2 + 10.92x_1x_2\right\} \\ -1 \le x_1x_2 \le 1 \end{cases}.$$

For the 2nd group:

$$\begin{cases} \min\left\{17.01 - 8.83x_1 - 8.34x_2 + 5.69x_1x_2\right\} \\ -1 \le x_1x_2 \le 1 \end{cases}.$$

Table 4.5 gives the minimal values for the number of impurities. It also shows the variance and the mean square error.

Table 4.5: Minimal values for the number of impurities.

Design	Optimal setting (x_1, x_2, x_3)	Number of impurities	Variance	*MSE*
Full design	$(-1 , 1 , 1)$	7.87	21.01	82.95
1st group	$(-1, 1, \forall x_3 \in R)$	6.17	30.69	68.76
2nd group	$(1, 1, \forall x_3 \in R)$	5.53	11.54	42.12

Interpretation

The analysis of variance permits depicting which main effects and interactions of the factors are significant to the number of impurities in the process. Once significant effects are detected, the regression models are fitted accordingly.

The results show that the factors x_1 and x_2 , which are respectively the reaction temperature and the catalyst concentration, have significant main effects and a significant interaction effect on the number of impurities. These results agree with those obtained in Chapter 1 and Chapter 2, where the analysis is based upon double orthogonal array design and combined array design respectively.

Furthermore, by analysing the data as a split-plot design, we obtain the best results for the number of impurities, compared to the results obtained in Chapter 1 and Chapter 2.

Application of the Quality Loss Function

The estimate of the expected loss is $\widehat{Q} = 25\left[\left(\widehat{\mu}\right)^2 + \widehat{\sigma^2}\right]$ The values obtained for $\widehat{\mu}$ and $\widehat{\sigma^2}$ are respectively 7.87 and 21.01. The corresponding quality loss is:

$\widehat{Q} = 25\left[\left(7.87\right)^2 + 21.01\right] = 2073.7$ monetary units.

The reduction of the quality loss is $7476.25 - 2073.7 = 5402.6$ monetary units, for each unit of chemical product.

4.4. Conclusions

In this chapter we briefly give a general view of split-plot design and introduce split-plot design for robust experimentation. The main idea in split-plot analysis is that the error term is split into parts, one for the whole-plots and one for the sub-plots. We mention the three experimental arrangements which are of the most interest while considering the design and noise factors. Through an illustrative example, we analyse the arrangement where the noise factors form the whole-plots and the design factors form the split-plots. The results show that split-plot analysis can be used to detect which factors influence the response of interest.

Chapter 5: Conclusions

In this thesis, we provide an overview of different approaches used in literature to conduct robust design and analyse the data obtained from this design. The general assumption in robust design is that there are two types of factors: the design factors which are controlled by the experimenter, and the noise factors which are difficult to control. As the noise factors are responsible for the variability which affects the quality characteristic of the process, the aim of robust design methodology is to achieve insensitivity to noise factors. This consists of bringing the quality characteristic to the target, while simultaneously minimizing its variance and the cost of the process.

The methods we present in this thesis are the Taguchi approach, the combined array design and the split-plot design for robust experimentation. The Taguchi approach consists of a double orthogonal array design, one for the design factors and one for the noise factors. Taguchi proposes the data analysis based upon the Signal-to-Noise Ratio. The combined array design puts both types of factors in one design. This design permits the analysis of the interactions between the designs and the noise factors, and reduces the number of runs required to conduct an experiment. The data analysis consists of adjusting 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 one for its variance. The split- plot design assigns noise factors in the whole plots and control factors in the sub-plots. The error term is formed of two terms.

We illustrate the Taguchi approach, the combined array design and the split-plot design with an application of a chemical process. The aim of the application is to conduct a designed experiment in order to obtain the operating conditions that improve the process. The improvement consists of the reduction of the number of impurities in the product, and the minimization of the cost. We use the quality loss function in order to evaluate the cost caused by the deviation of the quality characteristic from the target. Even though the optimal operating conditions obtained by data analysis based upon the Taguchi approach and the combined array design are slightly different, the common conclusion is that they reduce the number of impurities in the product and the cost of the process. The economic impact of conducting designed experiments is verified by using the Quality Loss Function.

By analysing the data as a split-plot design experiment, the results which are obtained agree with those of the Taguchi approach and the combined array design. These are that the factors x_1 and x_2, which are respectively the reaction temperature and the catalyst concentration, have significant main effects and a significant interaction effect on the number of impurities.

We expose the integrated method of parameter design and tolerance design. The parameter design methodology is built on engineering and statistical ideas, and aims at improving a system by making its performance insensitive to noise factors. This is attained by obtaining the optimal operating conditions in terms of the combination of the levels of the design factors. The tolerance design is introduced in the process when the variability of the quality characteristic is still large. A tighter tolerance around the nominal value results in a smaller transmitted variation, but incurs an extra cost. The method incorporates parameter design and tolerance design into a single stage of design optimization. We give an illustrative application of minimization of the cost in the production of polyamide resin.

Conclusiones

En esta tesina, proporcionamos un panorama de los diferentes enfoques usados en literatura para realizar un diseño robusto y analizar los datos obtenidos de este diseño. El supuesto general en diseño robusto es la presencia de dos tipos de factores: los factores de diseño que son controlados por el experimentador, y factores de ruido que son difíciles de controlar. Como los factores de ruido son responsables de la variabilidad que afecta a la característica de calidad del proceso, la meta de la metodología del diseño robusto es alcanzar insensibilidad a los factores de ruido. Esto consiste en obtener el valor objetivo de la característica de calidad, y minimizar simultáneamente la varianza y el costo de proceso.

Los métodos que presentamos en esta tesina son el enfoque de Taguchi, el diseño arreglo combinado y el diseño de parcelas divididas para experimentación robusta. El enfoque de Taguchi consiste en un diseño doble arreglo ortogonal, uno para factores de diseño y otro para factores de ruido. Taguchi propone el análisis de datos basado en la Razón Señal a Ruido. El diseño arreglo combinado considera ambos tipos de factores en un diseño. Este diseño permite el análisis de interacciones entre factores de diseño y de ruido, y reduce el número de tratamientos requeridos para realizar un experimento. El análisis de datos consiste en ajustar un modelo de regresión en términos de factores de diseño y factores de ruido. Del modelo ajustado, se obtienen dos superficies de respuesta, una para la media de la característica de calidad, y otra para la varianza. El diseño de parcelas divididas asigna los factores de ruido en toda la parcela y los de control a sub-parcelas. El término de error se divide en dos términos.

Ilustramos el enfoque de Taguchi, el diseño arreglo combinado y el diseño a parcelas divididas con un ejemplo de proceso químico. El objetivo de esta aplicación es realizar un experimento para obtener las condiciones de funcionamiento que mejoran el proceso. La mejora consiste en la reducción del número de impurezas en el producto, y minimización del costo. Utilizamos la función de pérdida de calidad para evaluar el costo causado por la desviación de la característica de calidad de su valor objetivo. Aunque las condiciones óptimas de funcionamiento obtenidas por el análisis de datos basado en enfoque de Taguchi sean ligeramente diferentes de las obtenidas en diseño arreglo combinado, la conclusión común es que reducen el número de impurezas en el producto y el costo del proceso. El impacto económico de realizar experimentos está verificado usando la función de pérdida de calidad.

Analizando los datos como diseño a parcelas divididas, los resultados permiten la misma conclusión que la obtenida en enfoque de Taguchi y en el diseño arreglo combinado. Esto es, los factores x_1 y x_2, que son respectivamente la temperatura de reacción y la concentración del catalizador, tienen efectos principales y efecto de interacciones significativos sobre el número de impurezas.

Exponemos el método integrado de diseño de parámetro y diseño de tolerancia. La metodología de diseño de parámetro está basada en ideas de ingeniería y estadística, y su objetivo es mejorar el sistema, haciendo su funcionamiento insensible a los factores de ruido. Esto se logra obteniendo las condiciones de funcionamiento óptimas en términos de combinación de los niveles de factores de diseño. El diseño de tolerancia se introduce en el proceso cuando la variabilidad de la característica de calidad es todavía grande. Una tolerancia más pequeña alrededor del valor nominal da lugar a una variación más pequeña, pero causa un costo adicional. El método integrado de diseño de parámetro y de tolerancia da lugar a un problema de optimización. Damos una aplicación ilustrativa de la minimización de costo en la producción de la poliamida resina.

References

[01] Alessandra Giovagnoli, Daniele Romano, (2008). *Robust Design via Simulation Experiments: A Modified Dual Response Surface Approach*. Quality and Reliability Engineering International, Vol. 24, 401-416.

[02] André I. Khuri, (2006). *Response Surface Methodology and Related Topics.* World Scientific Publishing Co. Pte. Ltd.

[03] André I. Khuri, Siuli Mukhopadhyay (2010). *Response Surface Methodology.* John Wiley & Son, Inc. WIREs Computational Statistics, Vol. 2, 128-149.

[04] Daniele Romano, Marco Varetto, Grazia Vicario, (2004). *Multiresponse Robust Design: A General Framework Based on Combined Array*. Journal of Quality Technology, Vol. 36, N°. 1, 27-37.

[05] Dennis K. J. Lin, Wanzhu Tu, (1995). *Dual Response Surface Optimization.* Journal of Quality Technology, Vol. 27, N°. 1, 34-39.

[06] Derek B., Randy R. S., (2003). *Fractional Factorial Split-Plot Designs for Robust Parameter Experiments*. Technometrics. Vol. 45, N°. 1, 80- 89.

[07] Domínguez D. J., (2006). *Optimización Simultánea para la Mejora Continua y Reducción de Costos en Procesos.* Ingeniería y Ciencias. Vol. 2, N°. 4, 145-162.

[08] Domínguez D. J., (2007). *Optimización de Costos en la Experimentación Industrial.* Revista de Matemática: Teoría y Aplicaciones. Vol. 14, N°. 1, 193-201.

[09] Domínguez D. J., Ernesto Barrios Zamudio, (2008). *Revisión de Métodos en Optimización Respuesta Dual.* To appear.

[10] Douglas C. Montgomery, (2001). *Design and Analysis of Experiments.* Wiley.

[11] Genichi Taguchi, (1988). *Introduction to Quality Engineering. Designing Quality into Products and Processes.* Asian Productivity Organization.

[12] George E. P. Box, Stephen Jones, (1992). *Designing Products That Are Robust To The Environment.* Total Quality Management, Vol. 3, N°. 3, 265-282.

[13] George E. P. Box, Stephen Jones, (1988). *Signal-to-Noise ratios, Performance, Criteria, and Transformations.* Technometrics, Vol. 30, 1-17.

[14] George E. P. Box, Stephen Jones, (1992). *Split- Plot Designs for Robust Product Experimentation.* Journal of Applied Statistics, Vol. 19, N°. 1, 3-26.

[15] Guillermo Miró-Quesada, Enrique del Castillo, (2004). *Two Approaches for Improving the Dual Response Method in Robust Parameter Design.* Journal of Quality Technology, Vol. 36, N°. 2, 154- 168.

[16] Jeff Wu C. F., Michael Hamada, (200). *Experiments: Planning, Analysis, and Parameter Design Optimization.* John Wiley & Son, Inc.

[17] Jennifer D. Letsinger, Raymond H. Myers, Marvin Lentner, (1996). *Response Surface Methods for Bi- Randomization Structures.* Journal of Quality Technology, Vol. 28, N°. 4, 381- 397.

[18] Joe Meng, Chuck Zhang, Ben Wang, (2010). *Integrated Robust Parameter Design and Tolerance Design.* Int. J. Industrial and Systems Engineering, Vol.5, N°. 2, 159-189.

[19] John S. Lawson, (1990). *Improving a Chemical Process Through Use Of a Designed Experiment.* Quality Engineering, Vol. 3, N°. 2, 215-235.

[20] Klaus Hinkelman, Oscar Kempthorne, (2007). *Design and Analysis of Experiments, Vol. I.* John Wiley & Sons.

[21] Klaus Hinkelman, Oscar Kempthorne, (2005). *Design and Analysis of Experiments, Vol. II.* John Wiley & Sons.

[22] Kunert J., Auer C., Erdbrügge M., Ewers R., (2007). *An Experiment to Compare Taguchi´s Product Array and the Combined Array.* Journal of Quality Technology, Vol. 39, N°. 1, 17-34.

[23] Martin Arvidsson, Ida Gremyr, (2008). *Principles of Robust Design Methodology*. Quality and Reliability Engineering International, Vol. 24, 23- 35.

[24] Myers R. H., Carter W. H., (1973). *Response Surface Techniques for Dual Response Systems*. Technometrics, Vol. 15, 301-317.

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

[26] Myers R. H., Khuri A. I., Vining G. G., (1992). *Response Surface Alternatives to the Taguchi Robust Parameter Design Approach*. The American Statistician, Vol. 46, 131- 139.

[27] Phillip J. Ross, (1988). *Taguchi Techniques for Quality Engineering. Loss Function, Orthogonal Experiments, Parameter and Tolerance Design*. McGraw- Hill Book Company.

[28] Raymond H. Myers, Douglas C. Montgomery, (1995). *Response Surface Methodology. Process and Product Optimization Using Designed Experiments*. John Wiley & Sons.

[29] Scott M. Kowalski, Peter A. Parker, G. Geoffrey Vining, (2007). *Tutorial: Industrial Split-plot Experiments*. Quality Engineering, Vol. 19, 1-15.

[30] Soren Bisgaard, (2000). *The Design and Analysis of* $2^{k-p}\times 2^{q-r}$ *Split Plot Experiments*. Journal of Quality Technology, Vol. 32, N°. 1, 39-56.

[31] Tang L. C., Xu K., (2002). *A Unified Approach for Dual Response Surface Optimization*. Journal of Quality Technology, Vol. 34, N°. 4, 437-447.

[32] Timothy J. Robinson, Connie M. Borror, Raymond H. Myers, (2004). *Robust Parameter Design: A Review*. Quality and Reliability Engineering International, Vol. 20, 81-101.

[33] Vijayan N. Nair, (1992). *Taguchi´s Parameter Design: A Panel Discussion.* Technometrics, Vol. 34, N°2, 128-161.

[34] William Li, C. F. J. Wu, (1999). *An Integrated Method of Parameter Design and Tolerance Design.* Quality Engineering, Vol. 11, 417-425.

[35] William Y. Fowlkes, Clyde M. Creveling, (1995). *Engineering Methods for Robust Product Design. Using Taguchi Methods in Technology and Product Development.* Addison-Wesley Publishing Company.

[36] Young J. K., Byung R. C., (2000). *Economic Considerations on Parameter Design.* Quality and Reliability Engineering International, Vol. 16, 501-514.

[37] Maghsooldoo S., (1990). *The Exact Relationship of Taguchi´s Signal-to-Noise Ratio to His Quality Loss Function.* Journal of Quality Technology, Vol. 22, N°. 1, 57-67.

[38] Mae Fah Luang University , School of Agro-industry.
http://www.mfu.ac.th/2009mfu_eng/U_GraduateBulletin2008/agro.pdf

Annexes

This part of Annexes contains tables and graph of the data used as applications in this thesis. Annexes are presented by chapter.

Chapter 1: Double Orthogonal Array Design

Annex 1.1: 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	

Annex 1.2: Sixty runs of the chemical process.

				Z1→	-1	-1	1	Mean	Stand. Dev.	Variance	SNR
Exp.	X1	X2	X3	Z2-	-1	1	1				
1	-1	-1	0	57.81	37.29	42.87	47.07	46.26	8.68	75.34	-33.4173
2	1	-1	0	24.89	4.35	8.23	14.69	13.04	8.98	80.60	-23.6265
3	-1	1	0	13.21	9.51	10.1	11.19	11.00	1.63	2.65	-20.9006
4	1	1	0	13.29	9.15	10.3	11.23	10.99	1.75	3.07	-20.9039
5	-1	0	-1	27.71	20.24	22.28	24.23	23.62	3.18	10.11	-27.5224
6	1	0	-1	11.4	4.48	5.44	8.23	7.39	3.11	9.69	-17.9126
7	-1	0	1	30.65	18.4	20.24	24.45	23.44	5.44	29.55	-27.5691
8	1	0	1	14.94	2.29	4.3	8.49	7.51	5.59	31.24	-19.0175
9	0	-1	-1	42.68	22.42	21.64	30.3	29.26	9.76	95.34	-29.6739
10	0	1	-1	13.56	10.08	9.85	11.38	11.22	1.70	2.89	-21.0722
11	0	-1	1	50.6	13.19	18.84	30.97	28.40	16.55	274.06	-30.0523
12	0	1	1	15.21	7.44	9.78	11.82	11.06	3.29	10.85	-21.1566
13	0	0	0	19.62	12.29	13.14	14.54	14.90	3.28	10.77	-23.6176
14	0	0	0	20.6	11.49	12.06	13.49	14.41	4.21	17.74	-23.443
15	0	0	0	20.15	12.2	14.06	13.89	15.08	3.49	12.15	-23.7359

Annex 1.3: ANOVA for the Mean.

Analysis of Variance for Mean

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
A:x1	534.482	1	534.482	2771.08	0.0000
B:x2	660.48	1	660.48	3424.33	0.0000
C:x3	0.1458	1	0.1458	0.76	0.4244
AA	0.980878	1	0.980878	5.09	0.0738
AB	275.726	1	275.726	1429.53	0.0000
AC	0.0225	1	0.0225	0.12	0.7466
BB	92.6927	1	92.6927	480.58	0.0000
BC	0.1225	1	0.1225	0.64	0.4616
CC	0.116878	1	0.116878	0.61	0.4715
Total error	0.964392	5	0.192878		
Total (corr.)	1564.87	14			

R-squared = 99.9384 %

R-squared (adjusted for d.f.) = 99.8274 %

Annex 1.4: ANOVA for the Standard deviation.

Analysis of Variance for Standard Deviation

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
A:x1	0.0323215	1	0.0323215	0.02	0.8949
B:x2	158.44	1	158.44	94.71	0.0002
C:x3	21.5073	1	21.5073	12.86	0.0158
AA	3.33187	1	3.33187	1.99	0.2172
AB	0.00755161	1	0.00755161	0.00	0.9490
AC	0.012133	1	0.012133	0.01	0.9354
BB	23.9977	1	23.9977	14.35	0.0128
BC	6.75376	1	6.75376	4.04	0.1007
CC	9.67948	1	9.67948	5.79	0.0612
Total error	8.36417	5	1.67283		
Total (corr.)	232.448	14			

R-squared = 96.4017%

R-squared (adjusted for d.f.) = 89.9247%

Annex 1.5: ANOVA for the Variance.

Analysis of Variance for Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
A:x1	6.03781	1	6.03781	0.00	0.9533
B:x2	31989.2	1	31989.2	20.10	0.0065
C:x3	6479.15	1	6479.15	4.07	0.0997
AA	2196.56	1	2196.56	1.38	0.2930
AB	5.86124	1	5.86124	0.00	0.9540
AC	1.11514	1	1.11514	0.00	0.9799
BB	9698.31	1	9698.31	6.09	0.0566
BC	7290.0	1	7290.0	4.58	0.0853
CC	3544.1	1	3544.1	2.23	0.1959
Total error	7958.41	5	1591.68		
Total (corr.)	69563.2	14			

R-squared = 88.5594%
R-squared (adjusted for d.f.) = 67.9664%

Annex 1.6: ANOVA for the SNR.

Analysis of Variance for SNR

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
A:x1	97.6426	1	97.6426	51.36	0.0008
B:x2	133.961	1	133.961	70.47	0.0004
C:x3	0.325786	1	0.325786	0.17	0.6961
AA	1.7328	1	1.7328	0.91	0.3836
AB	23.9811	1	23.9811	12.61	0.0164
AC	0.279947	1	0.279947	0.15	0.7169
BB	11.9404	1	11.9404	6.28	0.0541
BC	0.021609	1	0.021609	0.01	0.9192
CC	0.0309946	1	0.0309946	0.02	0.9034
Total error	9.50538	5	1.90108		
Total (corr.)	280.177	14			

R-squared = 96.6074 %
R-squared (adjusted for d.f.) = 90.5006 %

Annex 1.7: Runs for the half fractional design.

Red coloured data: 1st group

Blue coloured data: 2nd group

Exp.	X1	X2	X3	Z1+ Z2-	-1 -1	-1 1	1 1
1	-1	-1	0	57.81	37.29	42.87	47.07
2	1	-1	0	24.89	4.35	8.23	14.69
3	-1	1	0	13.21	9.51	10.1	11.19
4	1	1	0	13.29	9.15	10.3	11.23
5	-1	0	-1	27.71	20.24	22.28	24.23
6	1	0	-1	11.4	4.48	5.44	8.23
7	-1	0	1	30.65	18.4	20.24	24.45
8	1	0	1	14.94	2.29	4.3	8.49
9	0	-1	-1	42.68	22.42	21.64	30.3
10	0	1	-1	13.56	10.08	9.85	11.38
11	0	-1	1	50.6	13.19	18.84	30.97
12	0	1	1	15.21	7.44	9.78	11.82
13	0	0	0	19.62	12.29	13.14	14.54
14	0	0	0	20.6	11.49	12.06	13.49
15	0	0	0	20.15	12.2	14.06	13.89

Annex 1.8:

Results for the half fractional design (Mean- Standard deviation- Variance- SNR).

Exp	X1	X2	X3	Mean	Stand. Dev.	Variance	SNR
1	-1	-1	0	52.44	14.87	57.67	-34.4386
2	1	-1	0	6.29	12.10	7.53	-16.3677
3	-1	1	0	12.20	12.20	2.04	-21.7569
4	1	1	0	9.73	12.61	0.66	-19.7729
5	-1	0	-1	25.97	12.94	6.06	-28.3089
6	1	0	-1	4.96	13.36	0.46	-13.9501
7	-1	0	1	27.55	13.34	19.22	-28.8571
8	1	0	1	3.30	13.78	2.02	-10.7434
9	0	-1	-1	36.49	13.42	76.63	-31.3667
10	0	1	1	9.97	12.25	0.03	19.9701
11	0	-1	1	40.79	12.96	192.67	-32.4545
12	0	1	1	8.61	4.46	2.74	-18.7795
13	0	0	0	17.08	3.74	12.90	-24.7448
14	0	0	0	11.78	3.97	0.16	-21.4218
15	0	0	0	17.02	4.43	19.59	-24.7636

Exp.	X1	X2	X3	Mean	Stand. Dev.	Variance	SNR
1	-1	-1	0	40.08	3.95	15.57	-32.0795
2	1	-1	0	19.79	7.21	52.02	-26.2082
3	-1	1	0	9.81	0.42	0.17	-19.8329
4	1	1	0	12.26	1.46	2.12	-21.8004
5	-1	0	-1	21.26	1.44	2.08	-26.5613
6	1	0	-1	9.82	2.24	5.02	-19.9496
7	-1	0	1	19.32	1.30	1.69	-25.73
8	1	0	1	11.72	4.56	20.80	-21.6921
9	0	-1	-1	22.03	0.55	0.30	-26.8617
10	0	1	-1	12.47	1.54	2.38	-21.9504
11	0	-1	1	16.02	4.00	15.96	-24.2236
12	0	1	1	13.52	2.40	5.75	-22.6841
13	0	0	0	12.72	0.60	0.36	-22.0912
14	0	0	0	17.05	5.03	25.28	-24.8169
15	0	0	0	13.13	1.32	1.73	-22.387

Annex 1.9: ANOVA for the Mean.

1st group

Analysis of Variance for Mean

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
A:x1	1101.92	1	1101.92	67.99	0.0004
B:x2	1140.15	1	1140.15	70.35	0.0004
C:x3	1.01888	1	1.01888	0.06	0.8120
AA	12.2752	1	12.2752	0.76	0.4239
AB	476.876	1	476.876	29.43	0.0029
AC	2.63251	1	2.63251	0.16	0.7036
BB	165.521	1	165.521	10.21	0.0241
BC	7.98062	1	7.98062	0.49	0.5142
CC	14.4084	1	14.4084	0.89	0.3890
Total error	81.0311	5	16.2062		
Total (corr.)	3007.32	14			

R squared = 97.3055 %

R squared (adjusted for d.f.) = 92.4555 %

2nd group

Analysis of Variance for Mean

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
A:x1	170.063	1	170.063	8.72	0.0318
B:x2	310.815	1	310.815	15.94	0.0104
C:x3	3.13751	1	3.13751	0.16	0.7049
AA	30.0654	1	30.0654	1.54	0.2694
AB	129.334	1	129.334	6.63	0.0497
AC	3.6864	1	3.6864	0.19	0.6818
BB	41.0308	1	41.0308	2.10	0.2066
BC	12.4609	1	12.4609	0.64	0.4603
CC	9.72252	1	9.72252	0.50	0.5116
Total error	97.4894	5	19.4979		
Total (corr.)	808.984	14			

R-squared = 87.9492 %

R-squared (adjusted for d.f.) = 66.2576 %

Annex 1.10: ANOVA for the Standard deviation.

1st group

Analysis of Variance for Standard Deviation

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
A:x1	0,282489	1	0,282489	0,07	0,7953
B:x2	17,5205	1	17,5205	4,64	0,0837
C:x3	6,89244	1	6,89244	1,83	0,2344
AA	121,66	1	121,66	32,25	0,0024
AB	2,54498	1	2,54498	0,67	0,4488
AC	0,000180902	1	0,000180902	0,00	0,9947
BB	36,9148	1	36,9148	9,79	0,0260
BC	13,4172	1	13,4172	3,56	0,1180
CC	47,0515	1	47,0515	12,47	0,0167
Total error	18,8602	5	3,77204		
Total (corr.)	240,601	14			

R-squared = 92,1677 %
R-squared (adjusted for d.f.) = 78,0696 %

2nd group

Analysis of Variance for Standard Deviation

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
A:x1	8,7468	1	8,7468	2,06	0,2111
B:x2	12,2325	1	12,2325	2,87	0,1507
C:x3	5,2441	1	5,2441	1,23	0,3174
AA	1,34828	1	1,34828	0,32	0,5978
AB	1,24031	1	1,24031	0,29	0,6124
AC	1,5138	1	1,5138	0,36	0,5768
BB	0,42461	1	0,42461	0,10	0,7648
BC	1,67445	1	1,67445	0,39	0,5580
CC	1,04656	1	1,04656	0,25	0,6410
Total error	21,2748	5	4,25495		
Total (corr.)	54,9361	14			

R-squared = 61,2736 %
R-squared (adjusted for d.f.) = 0,0 %

Annex 1.11: ANOVA for the Variance.

1st group

Analysis of Variance for Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
A:x1	690.433	1	690.433	0.50	0.5096
B:x2	13533.1	1	13533.1	9.87	0.0256
C:x3	2226.85	1	2226.85	1.62	0.2585
AA	2791.07	1	2791.07	2.04	0.2130
AB	594.579	1	594.579	0.43	0.5393
AC	33.6748	1	33.6748	0.02	0.8816
BB	4164.27	1	4164.27	3.04	0.1418
BC	3210.58	1	3210.58	2.34	0.1865
CC	2047.17	1	2047.17	1.49	0.2762
Total error	6854.55	5	1370.91		
Total (corr.)	36662.3	14			

R-squared = 81.3035 %
R-squared (adjusted for d.f.) = 47.6499 %

2nd group

Analysis of Variance for Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
A:x1	456.801	1	456.801	2.73	0.1592
B:x2	674.097	1	674.097	4.03	0.1009
C:x3	148.056	1	148.056	0.89	0.3898
AA	85.986	1	85.986	0.51	0.5053
AB	297.632	1	297.632	1.78	0.2396
AC	65.326	1	65.326	0.39	0.5593
BB	45.8233	1	45.8233	0.27	0.6229
BC	37.7432	1	37.7432	0.23	0.6547
CC	158.328	1	158.328	0.95	0.3751
Total error	835.664	5	167.133		
Total (corr.)	2828.77	14			

R-squared = 70.4584 %
R-squared (adjusted for d.f.) = 17.2835 %

Annex 1.12: ANOVA for the SNR.

1st group

Analysis of Variance for SNR

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
A:x1	344.891	1	344.891	29.39	0.0029
B:x2	147.474	1	147.474	12.57	0.0165
C:x3	0.953097	1	0.953097	0.08	0.7871
AA	30.3835	1	30.3835	2.59	0.1685
AB	64.6971	1	64.6971	5.51	0.0657
AC	3.52482	1	3.52482	0.30	0.6072
BB	19.6893	1	19.6893	1.68	0.2518
BC	1.29778	1	1.29778	0.11	0.7530
CC	0.354659	1	0.354659	0.03	0.8688
Total error	58.6801	5	11.736		
Total (corr.)	675.946	14			

R-squared = 91.3188%
R-squared (adjusted for d.f.) = 75.6927 %

2nd group

Analysis of Variance for SNR

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
A:x1	26.4752	1	26.4752	5.48	0.0663
B:x2	66.7313	1	66.7313	13.81	0.0138
C:x3	0.123306	1	0.123306	0.03	0.8794
AA	1.90131	1	1.90131	0.39	0.5581
AB	15.3617	1	15.3617	3.18	0.1347
AC	1.65611	1	1.65611	0.34	0.5837
BB	5.0052	1	5.0052	1.04	0.3555
BC	2.84226	1	2.84226	0.59	0.4778
CC	0.408719	1	0.408719	0.08	0.7829
Total error	24.1671	5	4.83343		
Total (corr.)	144.642	14			

R-squared = 83.2917 %
R-squared (adjusted for d.f.) = 53.2169 %

Chapter 2: Combined Array Design

Annex 2.1: Sixty runs of the chemical process.

x_1	x_2	x_3	z_1	z_2	Impurity
-1	-1	0	1	-1	57.81
1	-1	0	1	-1	24.89
-1	1	0	1	-1	13.21
1	1	0	1	-1	13.29
-1	0	-1	1	-1	27.71
1	0	-1	1	-1	11.40
-1	0	1	1	-1	30.65
1	0	1	1	-1	14.94
0	-1	-1	1	-1	42.68
0	1	-1	1	-1	13.56
0	-1	1	1	-1	50.60
0	1	1	1	-1	15.21
0	0	0	1	-1	19.62
0	0	0	1	-1	20.60
0	0	0	1	-1	20.15
-1	-1	0	-1	-1	37.29
1	-1	0	-1	-1	4.35
-1	1	0	-1	-1	9.51
1	1	0	-1	-1	9.15
-1	0	-1	-1	-1	20.24
1	0	-1	-1	-1	4.48
-1	0	1	-1	-1	18.40
1	0	1	-1	-1	2.29
0	-1	-1	-1	-1	22.42
0	1	-1	-1	-1	10.08
0	-1	1	-1	-1	13.19
0	1	1	-1	-1	7.44
0	0	0	-1	-1	12.29
0	0	0	-1	-1	11.49
0	0	0	-1	-1	12.20

x_1	x_2	x_3	z_1	z_2	Impurity
-1	-1	0	-1	1	42.87
1	-1	0	-1	1	8.23
-1	1	0	-1	1	10.10
1	1	0	-1	1	10.30
-1	0	-1	-1	1	22.28
1	0	-1	-1	1	5.44
-1	0	1	-1	1	20.24
1	0	1	-1	1	4.30
0	-1	-1	-1	1	21.64
0	1	-1	-1	1	9.85
0	-1	1	-1	1	18.84
0	1	1	-1	1	9.78
0	0	0	-1	1	13.14
0	0	0	-1	1	12.06
0	0	0	-1	1	14.06
-1	-1	0	1	1	47.07
1	-1	0	1	1	14.69
-1	1	0	1	1	11.19
1	1	0	1	1	11.23
-1	0	-1	1	1	24.23
1	0	-1	1	1	8.23
-1	0	1	1	1	24.45
1	0	1	1	1	8.19
0	-1	-1	1	1	30.30
0	1	-1	1	1	11.38
0	-1	1	1	1	30.97
0	1	1	1	1	11.82
0	0	0	1	1	14.54
0	0	0	1	1	13.49
0	0	0	1	1	13.89

Annex 2.2: ANOVA for the number of impurities.

Analysis of Variance for Impurity

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	7688.99	17	452.294	33.40	0.0000
Residual	568.683	42	13.5401		

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
x1	2137.76	1	2137.76	157.88	0.0000
x2	2641.55	1	2641.55	195.09	0.0000
x3	0.580503	1	0.580503	0.04	0.8370
x1*x1	3.91875	1	3.91875	0.29	0.5934
x1*x2	1102.9	1	1102.9	81.45	0.0000
x1*x3	0.0885063	1	0.0885063	0.01	0.9359
x2*x2	371.372	1	371.372	27.43	0.0000
x2*x3	0.497025	1	0.497025	0.04	0.8490
x3*x3	0.459335	1	0.459335	0.03	0.8548
z1	915.254	1	915.254	67.60	0.0000
z2	86.496	1	86.496	6.39	0.0153
x1*z1	0.326028	1	0.326028	0.02	0.8774
x1*z2	0.0693781	1	0.0693781	0.01	0.9433
x2*z1	347.82	1	347.82	25.69	0.0000
x2*z2	33.661	1	33.661	2.49	0.1224
x3*z1	48.9803	1	48.9803	3.62	0.0641
x3*z2	0.664128	1	0.664128	0.05	0.8258
Residual	568.683	42	13.5401		
Total (corrected)	8257.68	59			

R-Squared = 93.1133 %
R-Squared (adjusted for d.f.) = 90.3258%
Standard Error of Est. = 3.67968

Annex 2.3: Coefficients of the regression model for the number of impurities.

Parameter	Estimate	Standard Error	Lower Limit	Upper Limit
CONSTANT	14.7942	1.06223	12.6505	16.9378
x1	-8.17344	0.650482	-9.48617	-6.86071
x2	-9.08563	0.650482	-10.3984	-7.7729
x3	-0.134688	0.650482	-1.44742	1.17804
x1*x1	0.515104	0.957484	-1.41718	2.44739
x1*x2	8.3025	0.919921	6.44602	10.159
x1*x3	0.074375	0.919921	-1.7821	1.93085
x2*x2	5.01448	0.957484	3.08219	6.94676
x2*x3	0.17625	0.919921	-1.68023	2.03273
x3*x3	0.176354	0.957484	-1.75593	2.10864
z1	3.90567	0.475045	2.94698	4.86435
z2	-1.20067	0.475045	-2.15935	-0.241985
x1*z1	0.100937	0.650482	-1.21179	1.41367
x1*z2	-0.0465625	0.650482	-1.35929	1.26617
x2*z1	-3.29688	0.650482	-4.6096	-1.98415
x2*z2	1.02562	0.650482	-0.287104	2.33835
x3*z1	1.23719	0.650482	-0.0755417	2.54992
x3*z2	-0.144062	0.650482	-1.45679	1.16867

Annex 2.4: Thirty runs of the 1st group.

x_1	x_2	x_3	z_1	z_2	Impurity
-1	-1	0	1	-1	57.81
1	1	0	1	1	47.07
1	0	-1	-1	-1	4.35
-1	0	1	-1	1	8.23
0	-1	-1	1	-1	13.21
0	1	1	1	1	11.19
0	0	0	-1	-1	9.15
0	0	0	-1	1	10.30
1	-1	0	1	-1	27.71
-1	0	-1	1	1	24.23
-1	0	1	-1	-1	4.48
1	0	1	-1	1	5.44
0	1	-1	1	-1	30.65
0	0	0	1	1	24.45
0	0	0	-1	-1	2.29

x_1	x_2	x_3	z_1	z_2	Impurity
-1	-1	0	-1	1	4.30
-1	1	0	1	-1	42.68
1	0	-1	1	1	30.30
1	0	1	-1	-1	10.08
0	-1	-1	-1	1	9.85
0	-1	1	1	-1	50.60
0	0	0	1	1	30.97
-1	-1	0	-1	-1	7.44
1	-1	0	-1	1	9.78
1	1	0	1	-1	19.62
-1	0	1	1	1	14.54
0	-1	-1	-1	-1	11.49
0	1	-1	-1	1	12.06
0	1	1	1	-1	20.15
0	0	0	1	1	13.89

Annex 2.5: ANOVA for the number of impurities for the 1st group.

Analysis of Variance for Impurity

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	5030.32	17	295.901	2.56	0.0513
Residual	1384.72	12	115.393		

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
x1	2.48334	1	2.48334	0.02	0.8858
x2	30.711	1	30.711	0.27	0.6153
x3	0.571702	1	0.571702	0.00	0.9450
x1*x1	4.80466	1	4.80466	0.04	0.8417
x1*x2	43.4327	1	43.4327	0.38	0.5510
x1*x3	19.9567	1	19.9567	0.17	0.6848
x2*x2	91.9801	1	91.9801	0.80	0.3895
x2*x3	315.086	1	315.086	2.73	0.1244
x3*x3	42.7203	1	42.7203	0.37	0.5542
z1	1981.21	1	1981.21	17.17	0.0014
z2	11.8568	1	11.8568	0.10	0.7541
x1*z1	87.2524	1	87.2524	0.76	0.4016
x1*z2	73.0399	1	73.0399	0.63	0.4417
x2*z1	4.40409	1	4.40409	0.04	0.8484
x2*z2	45.601	1	45.601	0.40	0.5414
x3*z1	3.5739	1	3.5739	0.03	0.8632
x3*z2	119.184	1	119.184	1.03	0.3295
Residual	1384.72	12	115.393		
Total (corrected)	6415.03	29			

R-Squared = 78.4145 %
R-Squared (adjusted for d.f.) = 47.835%
Standard Error of Est. = 10.7421

Annex 2.6:

Coefficients of the regression model for the number of impurities (1st group).

Parameter	Estimate	Standard Error	Lower Limit	Upper Limit
CONSTANT	15.4113	4.44697	5.72215	25.1004
x1	0.458484	3.12533	-6.35104	7.26801
x2	-1.83406	3.55513	-9.58004	5.91193
x3	0.222792	3.16522	-6.67365	7.11923
x1*x1	1.24697	6.11103	-12.0679	14.5618
x1*x2	2.84178	4.63204	-7.25058	12.9341
x1*x3	-3.44924	8.29411	-21.5206	14.6221
x2*x2	6.45176	7.22639	-9.29322	22.1967
x2*x3	-7.81076	4.72681	-18.1096	2.48811
x3*x3	-3.21087	5.27711	-14.7087	8.28699
z1	9.99451	2.41205	4.7391	15.2499
z2	-0.708852	2.21137	-5.52703	4.10933
x1*z1	-5.23149	6.01626	-18.3398	7.87684
x1*z2	3.32252	4.17616	-5.77657	12.4216
x2*z1	1.14735	5.87295	-11.6487	13.9434
x2*z2	3.33198	5.30035	-8.21652	14.8805
x3*z1	-1.07892	6.13067	-14.4365	12.2787
x3*z2	-4.16897	4.10214	-13.1068	4.76884

Annex 2.7: Thirty runs of the 2nd group.

x_1	x_2	x_3	z_1	z_2	Impurity
1	-1	0	-1	-1	37.29
-1	1	0	-1	1	42.87
-1	0	-1	1	-1	24.89
1	0	1	1	1	14.69
0	1	-1	-1	-1	9.51
0	-1	1	-1	1	10.10
0	0	0	1	-1	13.29
-1	-1	0	1	1	11.23
-1	1	0	-1	-1	20.24
1	1	0	-1	1	22.28
1	0	-1	1	-1	11.40
0	-1	-1	1	1	8.23
0	-1	1	-1	-1	18.40
0	1	1	-1	1	20.24
0	0	0	1	-1	14.94

x_1	x_2	x_3	z_1	z_2	Impurity
1	-1	0	1	1	8.49
1	1	0	-1	-1	22.42
-1	0	-1	-1	1	21.64
-1	0	1	1	-1	13.56
0	1	-1	1	1	11.38
0	1	1	-1	-1	13.19
0	0	0	-1	1	18.84
0	0	0	1	-1	15.21
-1	1	0	1	1	11.82
-1	0	-1	-1	-1	12.29
1	0	-1	-1	1	13.14
1	0	1	1	-1	20.60
0	-1	1	1	1	13.49
0	0	0	-1	-1	12.20
0	0	0	-1	1	14.06

Annex 2.8: ANOVA for the number of impurities for the 2nd group.

Analysis of Variance for Impurity

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	1417.48	17	83.3814	2.84	0.0356
Residual	351.721	12	29.3101		

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
x1	9.09139	1	9.09139	0.31	0.5878
x2	23.285	1	23.285	0.79	0.3903
x3	59.1249	1	59.1249	2.02	0.1810
x1*x1	170.413	1	170.413	5.81	0.0328
x1*x2	219.953	1	219.953	7.50	0.0180
x1*x3	169.878	1	169.878	5.80	0.0331
x2*x2	116.525	1	116.525	3.98	0.0694
x2*x3	135.198	1	135.198	4.61	0.0529
x3*x3	0.459461	1	0.459461	0.02	0.9024
z1	243.713	1	243.713	8.32	0.0137
z2	11.9218	1	11.9218	0.41	0.5356
x1*z1	136.834	1	136.834	4.67	0.0516
x1*z2	189.131	1	189.131	6.45	0.0259
x2*z1	326.777	1	326.777	11.15	0.0059
x2*z2	306.36	1	306.36	10.45	0.0072
x3*z1	114.025	1	114.025	3.89	0.0721
x3*z2	20.1533	1	20.1533	0.69	0.4232
Residual	351.721	12	29.3101		
Total (corrected)	1769.21	29			

R-Squared = 80.1198 %
R-Squared (adjusted for d.f.) = 51.9563 %
Standard Error of Est. = 5.41388

Annex 2.9:

Coefficients of the regression model for the number of impurities (2nd group).

Parameter	Estimate	Standard Error	Lower Limit	Upper Limit
CONSTANT	14.5195	2.24127	9.63618	19.4028
x1	0.831521	1.49302	-2.42151	4.08455
x2	-1.49919	1.682	-5.16396	2.16559
x3	2.14999	1.51377	-1.14824	5.44821
x1*x1	5.76109	2.38925	0.555343	10.9668
x1*x2	-8.58085	3.13237	-15.4057	-1.75598
x1*x3	5.83549	2.42392	0.554216	11.1168
x2*x2	-7.8745	3.94931	-16.4793	0.730331
x2*x3	-7.1	3.30583	-14.3028	0.102814
x3*x3	0.284014	2.26842	-4.65847	5.22649
z1	-3.13432	1.08696	-5.5026	-0.766041
z2	-0.711494	1.1156	-3.14219	1.7192
x1*z1	-4.70427	2.17722	-9.44804	0.0395104
x1*z2	-3.93018	1.54718	-7.3012	-0.559164
x2*z1	-10.2783	3.07826	-16.9853	-3.57137
x2*z2	6.17889	1.91118	2.01477	10.343
x3*z1	-4.39083	2.22615	-9.24122	0.459548
x3*z2	-1.27069	1.53241	-4.60952	2.06815

Chapter 3:

Integrated method of Parameter design and Tolerance design

Annex 3.1: Factors, design and response values.

Level	**Temperature**	**Agitation**	**Rate**	x_1	x_2	x_3
High	200	10.0	25	+1	+1	+1
Center	175	7.5	20	0	0	0
Low	150	5.0	15	–1	–1	–1

Standard order	x_1	x_2	x_3	y
1	–1	–1	0	53
2	+1	–1	0	58
3	–1	+1	0	59
4	+1	+1	0	56
5	–1	0	–1	64
6	+1	0	–1	45
7	–1	0	+1	35
8	+1	0	+1	60
9	0	–1	–1	59
10	0	+1	–1	64
11	0	–1	+1	53
12	0	+1	+1	65
13	0	0	0	65
14	0	0	0	59
15	0	0	0	62

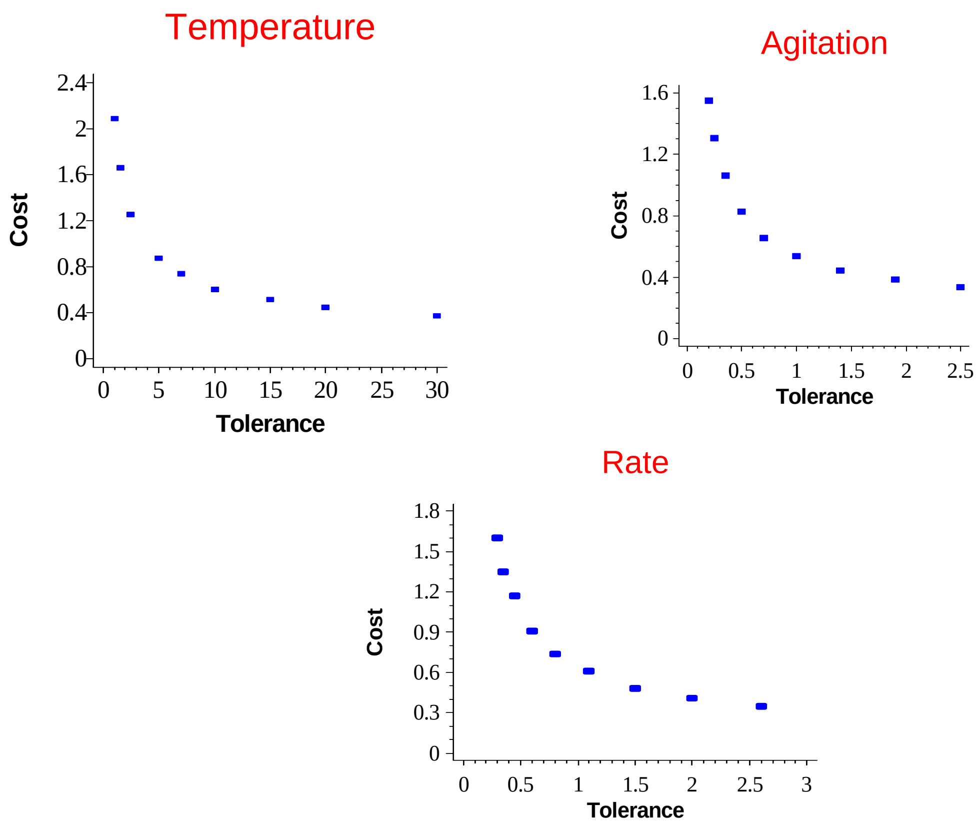
Annex 3.2: Analysis of variance for the viscosity of the polyamide resin.

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
A:x1	73.8174	1	73.8174	7.20	0.0436
B:x2	0.101077	1	0.101077	0.01	0.9248
C:x3	96.818	1	96.818	9.45	0.0277
AA	200.827	1	200.827	19.59	0.0069
AB	16.0	1	16.0	1.56	0.2668
AC	484.0	1	484.0	47.22	0.0010
BB	12.9808	1	12.9808	1.27	0.3115
BC	12.25	1	12.25	1.20	0.3241
CC	48.5192	1	48.5192	4.73	0.0816
Total error	51.25	5	10.25		
Total (corr.)	933.733	14			

Annex 3.3: Tolerances and costs.

Temperature		Agitation		Rate	
Tolerance	**Cost**	**Tolerance**	**Cost**	**Tolerance**	**Cost**
1	2.09	0.2	1.55	0.3	1.602
1.5	1.663	0.25	1.307	0.35	1.35
2.5	1.254	0.35	1.06	0.45	1.17
5	0.872	0.5	0.827	0.6	0.91
7	0.74	0.7	0.654	0.8	0.74
10	0.605	1	0.54	1.1	0.612
15	0.514	1.4	0.444	1.5	0.479
20	0.45	1.9	0.385	2	0.41
30	0.375	2.5	0.333	2.6	0.35

Cost vs Tolerance



Chapter 4: Split-Plot Design for Robust Experimentation

Annex 4.1: Sixty runs in Split-plot design structure.

WP	Z	x_1	x_2	x_3	Impurities
1	-1	-1	-1	0	57.8
1	-1	1	-1	0	24.9
1	-1	-1	1	0	13.2
1	-1	1	1	0	13.3
1	-1	-1	0	-1	27.7
1	-1	1	0	-1	11.4
1	-1	-1	0	1	30.7
1	-1	1	0	1	14.9
1	-1	0	-1	-1	42.7
1	-1	0	1	-1	13.6
1	-1	0	-1	1	50.6
1	-1	0	1	1	15.2
1	-1	0	0	0	19.6
1	-1	0	0	0	20.6
1	-1	0	0	0	20.2

WP	Z	x_1	x_2	x_3	Impurities
3	-1	-1	-1	0	42.87
3	-1	1	-1	0	8.23
3	-1	-1	1	0	10.1
3	-1	1	1	0	10.3
3	-1	-1	0	-1	22.28
3	-1	1	0	-1	5.44
3	-1	-1	0	1	20.24
3	-1	1	0	1	4.3
3	-1	0	-1	-1	21.64
3	-1	0	1	-1	9.85
3	-1	0	-1	1	18.84
3	-1	0	1	1	9.78
3	-1	0	0	0	13.14
3	-1	0	0	0	12.06
3	-1	0	0	0	14.06

WP	Z	x_1	x_2	x_3	Impurities
2	1	-1	-1	0	37.3
2	1	1	-1	0	4.35
2	1	-1	1	0	9.51
2	1	1	1	0	9.15
2	1	-1	0	-1	20.2
2	1	1	0	-1	4.48
2	1	-1	0	1	18.4
2	1	1	0	1	2.29
2	1	0	-1	-1	22.4
2	1	0	1	-1	10.1
2	1	0	-1	1	13.2
2	1	0	1	1	7.44
2	1	0	0	0	12.3
2	1	0	0	0	11.5
2	1	0	0	0	12.2

WP	Z	x_1	x_2	x_3	Impurities
4	1	-1	-1	0	47.07
4	1	1	-1	0	14.69
4	1	-1	1	0	11.19
4	1	1	1	0	11.23
4	1	-1	0	-1	24.23
4	1	1	0	-1	8.23
4	1	-1	0	1	24.45
4	1	1	0	1	8.49
4	1	0	-1	-1	30.3
4	1	0	1	-1	11.38
4	1	0	-1	1	30.97
4	1	0	1	1	11.82
4	1	0	0	0	14.54
4	1	0	0	0	13.49
4	1	0	0	0	13.89

Annex 4.2: Thirty runs of the 1st group.

WP	Z	x_1	x_2	x_3	Impurities
1	-1	-1	-1	0	57.8
1	-1	-1	1	0	13.2
1	-1	-1	0	-1	27.7
1	-1	-1	0	1	30.7
1	-1	0	-1	-1	42.7
1	-1	0	-1	1	50.6
1	-1	0	0	0	19.6
1	-1	0	0	0	20.2
2	1	1	-1	0	4.35
2	1	1	1	0	9.15
2	1	1	0	-1	4.48
2	1	1	0	1	2.29
2	1	0	1	-1	10.1
2	1	0	1	1	7.44
2	1	0	0	0	11.5

WP	Z	x_1	x_2	x_3	Impurities
3	-1	1	-1	0	8.2
3	-1	1	1	0	10
3	-1	1	0	-1	5.4
3	-1	1	0	1	4.3
3	-1	0	1	-1	9.9
3	-1	0	1	1	9.8
3	-1	0	0	0	12
4	1	-1	-1	0	47
4	1	-1	1	0	11
4	1	-1	0	-1	24
4	1	-1	0	1	24
4	1	0	-1	-1	30
4	1	0	-1	1	31
4	1	0	0	0	15
4	1	0	0	0	14

Annex 4.3:

Analysis of variance for the number of impurities considering the 1st group.

Source	DF	SS	Adj.SS	Adj. MS	F	P
Z	1	194.92	177.43	177.43	3.36	0.458
WP(Z)	2	3341.83	105.73	52.86	1.72	0.208
x1	1	57.50	476.04	476.04	15.51	0.001
x2	1	1251.78	1251.78	1251.78	40.78	0.000
x3	1	2.04	2.04	2.04	0.07	0.800
x1*x2	1	953.75	953.75	953.75	31.07	0.000
x1*x3	1	5.27	5.27	5.27	0.17	0.684
x2*x3	1	15.96	15.96	15.96	0.52	0.481
Z*x1	1	20.54	2.25	2.25	0.07	0.790
Z*x2	1	38.10	38.10	38.10	1.24	0.281
Z*x3	1	11.54	11.54	11.54	0.38	0.548
Error	17	521.79	521.79	30.69		
Total	29	6415.03				

Stand. error = 5.54019
R-squar. = 91.87%
R-squar.(adjust.) = 86.12%

Annex 4.4: Thirty runs of the 2nd group.

WP	Z	x_1	x_2	x_3	Impurities
1	-1	1	-1	0	24.89
1	-1	1	1	0	13.29
1	-1	1	0	-1	11.4
1	-1	1	0	1	14.94
1	-1	0	1	-1	13.56
1	-1	0	1	1	15.21
1	-1	0	0	0	20.6
2	1	-1	-1	0	37.29
2	1	-1	1	0	9.51
2	1	-1	0	-1	20.24
2	1	-1	0	1	18.4
2	1	0	-1	-1	22.42
2	1	0	-1	1	13.19
2	1	0	0	0	12.29
2	1	0	0	0	12.2

WP	Z	x_1	x_2	x_3	Impurities
3	-1	-1	-1	0	42.87
3	-1	-1	1	0	10.1
3	-1	-1	0	-1	22.28
3	-1	-1	0	1	20.24
3	-1	0	-1	-1	21.64
3	-1	0	-1	1	18.84
3	-1	0	0	0	13.14
3	-1	0	0	0	14.06
4	1	1	-1	0	14.69
4	1	1	1	0	11.23
4	1	1	0	-1	8.23
4	1	1	0	1	8.49
4	1	0	1	-1	11.38
4	1	0	1	1	11.82
4	1	0	0	0	13.49

Annex 4.5:

Analysis of variance for the number of impurities considering the 2nd group.

Source	DF	SS	Adj.SS	Adj. MS	F	P
Z	1	90.79	100.17	100.17	1.01	0.616
WP(Z)	2	239.23	197.64	98.82	8.56	0.003
x1	1	119.76	475.59	475.59	41.21	0.000
x2	1	792.17	792.17	792.17	68.64	0.000
x3	1	6.28	6.28	6.28	0.54	0.471
x1*x2	1	258.67	258.67	258.67	22.41	0.000
x1*x3	1	7.37	7.37	7.37	0.64	0.435
x2*x3	1	24.92	24.92	24.92	2.16	0.160
Z*x1	1	0.11	3.53	3.53	0.31	0.588
Z*x2	1	26.53	26.53	26.53	2.30	0.148
Z*x3	1	7.18	7.18	7.18	0.62	0.441
Error	17	196.20	196.20	11.54		
Total	29	1769.21				

Stand. error = 3.39721
R-squar. = 88.91%
R-squar.(adjust.) = 81.08%