© 1999-2000 Umetrics AB

Edition date: February 07 2000

© 2000 Learnways AB

Learnways manages and develops concepts for the design, production, marketing and distribution of web-based corporate education.

Learnways has developed a unique project method for the production and development of interactive educations. We have qualified competence within conceptualization, netbased pedagogics, multimedia productions, marketing and sales. This specific product has been developed in co-operation with Utero AB.

A web-based education from Learnways meets with the most critical customer needs on the market regarding competence development, offering the most cost- and time effective solution, the best contents, customized versions, continuous accessibility as well as relevant and interactive pedagogics.

Learnways was founded in 1997 and is based in Stockholm. The company is owned by ABB, Ericsson and Volvo, among others.

Contents

7 Factorial designs (Level 1)   59

8 Factorial designs (Level 2)   67

9 Factorial designs (Level 3)   77

10 Analysis of factorial designs (Level 1)   85

11 Analysis of factorial designs (Level 2)   93

12 Analysis of factorial designs (Level 3)   103

13 Experimental objective: Screening (Level 2)   113

20 Multi-level qualitative factors (Level 3) 239

21 Taguchi approach to robust design (Level 3) 247

22 Models of variability: Distributions (Level 3) 265

23 Analysis of variance, ANOVA (Level 3)   273

24 PLS (Level 3)   279

Statistical Appendix   293

References 321

Index 325

0 Preface

Why experiments?

How can we find the optimum? Which combination of our factors will give the best car engine, at lowest possible cost, using lowest possible fuel consumption, and producing a minimum of pollution? These and other related questions are common everywhere in business, research, and industry. In research, development, and production, often half of the available experimental resources are spent on solving optimization problems. With the rapidly increasing costs of experiments, it is essential that these questions are answered with as few experiments as possible. Design of Experiments, DOE, is used for this purpose – to ensure that the selected experiments are maximally informative.

Why Design of Experiments (DOE) is used

- Development of new products and processes
- Enhancement of existing products and processes
- Optimization of quality and performance of a product
- Optimization of an existing manufacturing procedure
- Screening of important factors
- Minimization of production costs and pollution
- Robustness testing of products and processes

Sectors where DOE is used

- Chemical industry
- Polymer industry
- Car manufacturing industry
- Pharmaceutical industry
- Food and dairy industry
- Pulp and paper industry
- Steel and mining industry
- Plastics and paints industry
- Telecom industry

- Marketing; Conjoint analysis

Three primary experimental objectives

Basically, there are three main types of problems to which DOE is applicable. In this book, these problem areas will be called *experimental objectives*. The first experimental objective is screening. Screening is used to uncover the most influential factors, and to determine in which ranges these should be investigated. This is a rather uncomplicated question, and therefore screening designs use few experiments in relation to the number of factors. The second experimental objective is optimization. Now, the interest lies in defining which approved combination of the important factors will result in optimal operating conditions. Since optimization is more complex than screening, optimization designs demand more experiments per factor. The third experimental objective is robustness testing. Here, one wants to determine how sensitive a product or production procedure is to small changes in the factor settings. Such small changes usually correspond to fluctuations in the factors occurring during "a bad day" in the production, or the customer not following product usage instructions.

The three primary experimental objectives

Screening
Which factors are most influential?
What are their appropriate ranges?

Optimization
How can we find the optimum?
Is there a unique optimum, or is a compromise necessary to meet conflicting demands on the responses?

Robustness testing
How should we adjust our factors to guarantee robustness?
Do we have to change our product specifications prior to claiming robustness?

The "intuitive" approach to experimental work

We may ask ourselves – *how is experimental work traditionally done*? Let us consider the optimization of a product or a process. This involves regulating the important factors so that the result becomes optimal. Usually this is done by changing the value of *one* separate factor at a time until no further improvement is accomplished. This is called the COST approach, and represents the intuitive way of performing experiments. An illustration to the COST-approach is given in Figure 0.1. Unfortunately, this is an inefficient approach. At the beginning of this century it was proven that changing one factor at a time does *not* necessarily give information about the optimum. This is particularly true when there are interactions among the factors. Then the COST-approach gets trapped, usually far from the real optimum. The problem is that the experimenter perceives that the optimum *has* been reached, simply because changing one factor at a time does not lead to any further improvement in the result. Changing one separate factor at a time (COST) does not lead to the real optimum, and gives different implications with different starting points, see Figure 0.1. The two factors, x_1 and x_2, are varied one at a time, making it difficult to reach the optimum.
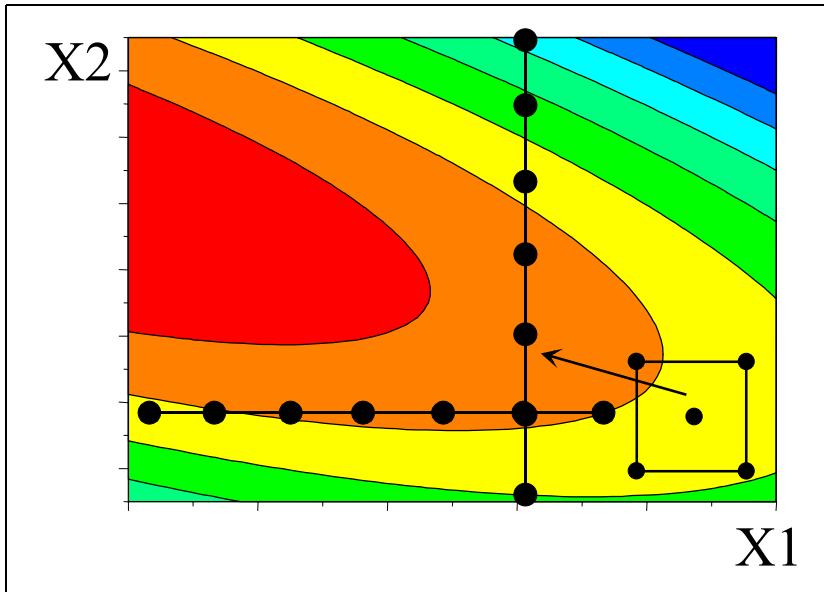
Figure 0.1: *An illustration of the COST approach and the DOE approach. Changing all factors at the same time according to a statistical experimental design (lower right corner) provides better information about the location of the optimum than the COST approach. Since the experiments here are laid out in a rectangular fashion a direction is obtained in which a better result is likely to be found*

A better approach – DOE

If we are not to use the COST-approach, what do we do instead? The solution is to construct a carefully selected *set* of experiments, in which *all* relevant factors are varied simultaneously. This is called statistical experimental design, or, design of experiments. Such a set of experiments usually does not contain more than 10-20 runs, but this number may be tailored to meet specific demands. In Figure 0.1 it is shown how two factors are studied at the same time using a small experimental design. Since the experiments are distributed in a rectangular fashion, a direction will be obtained in which a better result is likely to be found. The actual data analysis of the performed experiments will identify the optimal conditions, and reveal which factors influence the result. In other words, DOE provides a reliable basis for decision-making. DOE thus provides a framework for changing all important factors systematically, and with a limited number of experiments.

Overview of DOE

Prior to doing any experiments, the experimenter has to specify some input conditions (see Figure 0.2); the number of factors and their ranges, the number of responses, and the experimental objective. Then the experimental design is created, and its experiments are carried out, either in parallel, or one after another. Each experiment gives some results, that is, values of the response variables. Thereafter, these data are analyzed by regression analysis. This gives a model relating the changes in the factors to the changes in the responses. The model will indicate which factors are important, and how they combine in influencing the responses. The modelling results may also be converted into response

contour plots, so called maps, which are used to clarify where the best operating conditions are to be expected.

	1. Define Factors	2. Define Responses
	3. Create Design	4. Make Model
	5. Evaluate Model	6. Interpret and Use Model (make decisions)
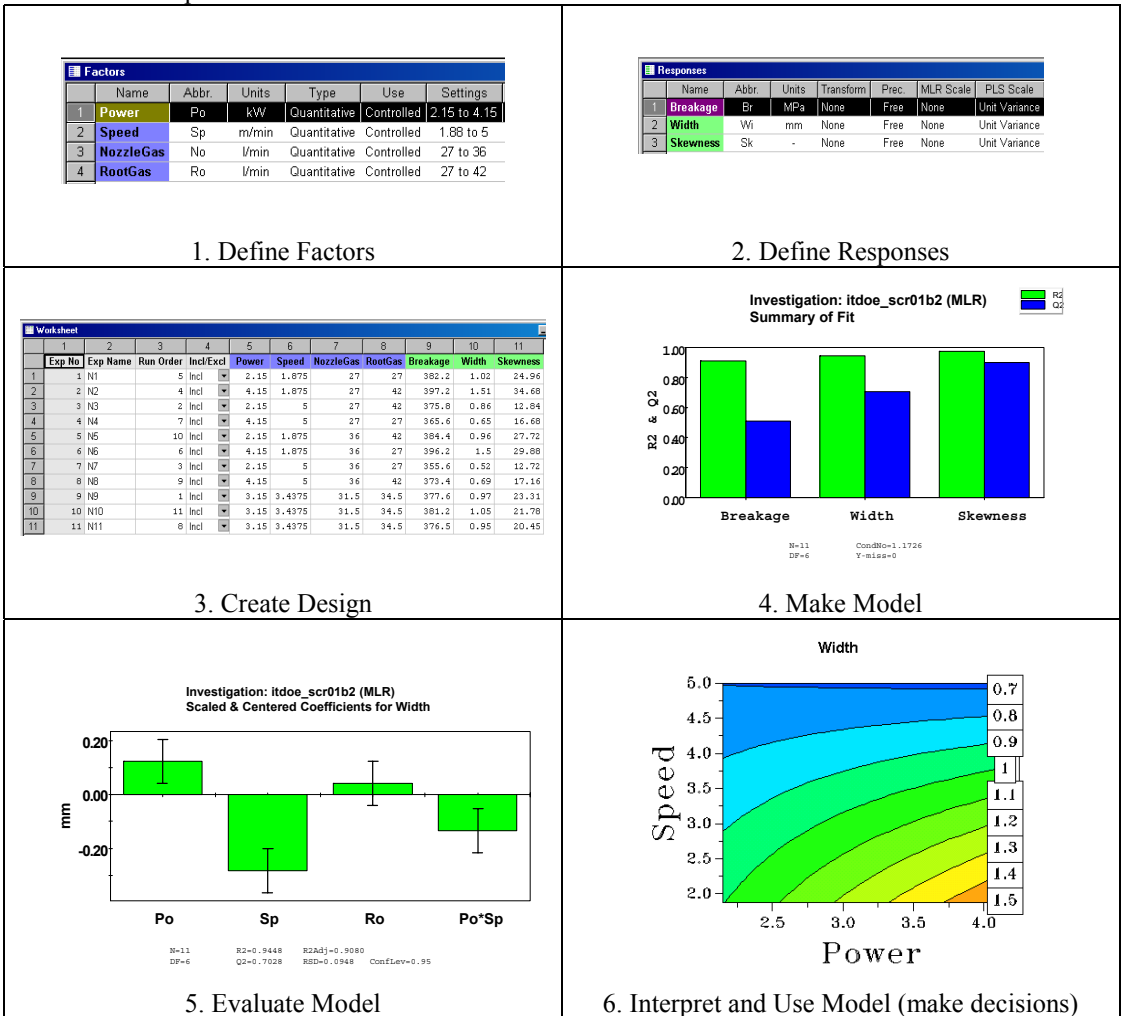
Figure 0.2: *An overview of some common steps in DOE.*

The objectives of the DOE course

The objectives of this DOE course can be summarized as four important points. Item 1: *How to do experiments efficiently*. The course will teach how to specify a problem definition and from that plan a series of experiments, which spans the experimental domain and is informative enough to answer the raised questions. Item 2*: How to analyze the data*. The mere fact that we have done several experiments does *not* guarantee that we know anything about the problem. Our raw data must be processed, analyzed, to extract the information in the data. This is accomplished with good statistical tools. Item 3: *How to interpret the results*. The data analysis will result in a mathematical model describing the relationships between factors and responses. We will place a lot of emphasis on how to

interpret such a model, that is, how to understand its message about the state of the investigated problem. This kind of interpretation is greatly facilitated by the use of graphical tools. Item 4: *How to put model interpretation into concrete action*. The aim of the last point is to encourage the user to convert the model interpretation into concrete plans of what to do next experimentally. In the ideal case, all problems are solved as a result of the first experimental series, and there is really no need for more experiments other than a couple of verifying trials. Often, however, more experiments are needed. It is then important to be able to decide how many, and which additional experiments, are most suited in each situation. In conclusion, the overall objective of the course is to make the student confident enough to master the four items just mentioned.

Organization of the book

This course book consists of 24 chapters, divided into 3 parts. The first part, i.e., chapters 1-12, provides a thorough introduction to the basic principles of DOE, by focusing on two-level full factorial designs. The second part, i.e., chapters 13-17, is application oriented and presents typical examples of screening, optimization, and robustness testing studies. The third part, chapters 18-24, is intended to equip the reader with insight into interesting additional topics and fundamental statistical concepts and methods. In the course, three levels of theory will be defined. The first level, Level 1, is aimed primarily for persons wishing to get a first impression of DOE. Level 1 runs only in Part 1. The second level, Level 2, is intended for users with some experience, but who may still need supervision in their daily work. This level runs in Parts 1 and 2. The third level, Level 3, is intended for advanced users working independently with DOE. This level runs in Parts 1, 2, and 3. The level classification is indicated in the title of each chapter.

Summary

In this chapter, we have highlighted three important experimental objectives for which DOE is useful. These are screening, optimization and robustness testing. In essence, DOE implies that a series of representative experiments is carried out, in which all factors are varied systematically at the same time. Thus, DOE corresponds to a rigorous framework for planning experiments, performing experiments, analyzing data, evaluating the resulting model, and converting modelling results to informative maps of the explored system.

Part I

Chapters 1-12

Basic principles of DOE

1 Introduction (Level 1)

Objective

In this chapter, we shall discuss where and how design of experiments, DOE, is used by industry today. Regardless of whether the experimental work takes place in the laboratory, the pilot plant, or the full-scale production plant, design of experiments is useful for three primary experimental objectives, screening, optimization and robustness testing. We shall discuss these three objectives and introduce one general example corresponding to each objective. These three examples will accompany us throughout the course.

We shall also describe another industrial application, the CakeMix application, which will help us to highlight some of the key elements involved in DOE. This example shows how changes in some important factors, that is, ingredients, can be linked to the changes in the response, that is, taste. In this way, an understanding is gained of how one must proceed to modify the amounts of the ingredients to improve the taste. The aim of this chapter is also to overview some of the most commonly employed DOE design families and point out when they are meaningful. The chapter ends with some arguments emphasizing the main benefits of adhering to the DOE methodology.

When and where DOE is useful

Design of experiments, DOE, is used in many industrial sectors, for instance, in the development and optimization of manufacturing processes. Typical examples are the production of wafers in the electronics industry, the manufacturing of engines in the car industry, and the synthesis of compounds in the pharmaceutical industry. Another main type of DOE-application is the optimization of analytical instruments. Many applications are found in the scientific literature describing the optimization of spectrophotometers and chromatographic equipment.

Usually, however, an experimenter does not jump directly into an optimization problem; rather initial screening experimental designs are used in order to locate the most fruitful part of the experimental region in question. Other main types of application where DOE is useful is robustness testing and mixture design. The key feature of the latter application type is that all factors sum to 100%.

Areas where DOE is used in industrial research, development and production:

- optimization of manufacturing processes

- optimization of analytical instruments

- screening and identification of important factors

- robustness testing of methods
- robustness testing of products
- formulation experiments

What is DOE?

One question which we might ask ourselves at this stage is *what is design of experiments*? DOE involves making a set of experiments representative with regards to a given question. The way to do this is, of course, problem dependent, and in reality the shape and complexity of a statistical experimental design may vary considerably. A common approach in DOE is to define an interesting standard reference experiment and then perform new, representative experiments around it (see Figure 1.1). These new experiments are laid out in a symmetrical fashion around the standard reference experiment. Hence, the standard reference experiment is usually called the *center-point*.
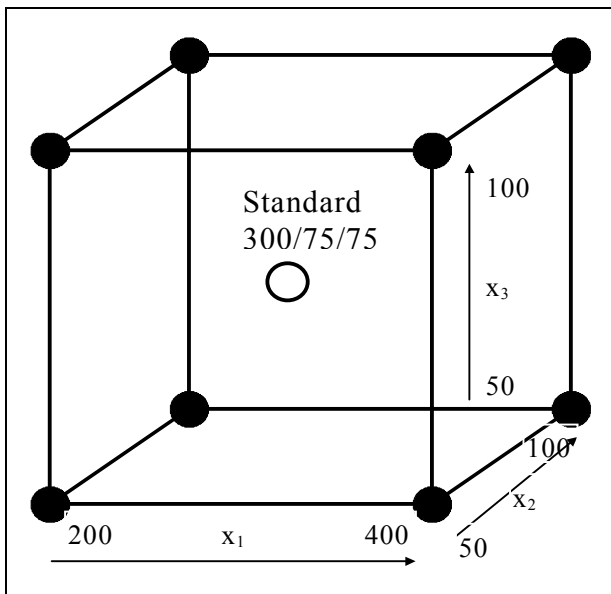
Figure 1.1: A symmetrical distribution of experimental points around a center-point experiment.

In the given illustration, the standard operating condition was used as the center-point. It prescribed that the first factor (x_1) should be set at the value 300, the second factor (x_2) at 75, and the third factor (x_3) at 75. In the next step, these three factors were varied according to the cubic pattern shown in Figure 1.1. This cubic pattern arises because the three factors are varied systematically around the center-point experiment. Thus, the first factor, x_1, is tested at a level slightly below the center-point, the value 200, and at a level slightly above the center-point, the value 400. A similar reasoning applies to factors x_2 and x_3.

Moreover, at a later stage in the experimental process, for instance, at an optimization step, already performed screening experiments may be used to predict a suitable reference experiment for an optimization design.

In the next three sections, we will introduce three representative DOE applications, which will accompany us in the course.

General Example 1: Screening

Screening is used at the beginning of the experimental procedure. The objective is (i) to explore many factors in order to reveal whether they have an influence on the responses, and (ii) to identify their appropriate ranges. Consider the laser welding material displayed in Figure 1.2. This is a cross-section of a plate heat-exchanger developed and manufactured by Alfa Laval Thermal.
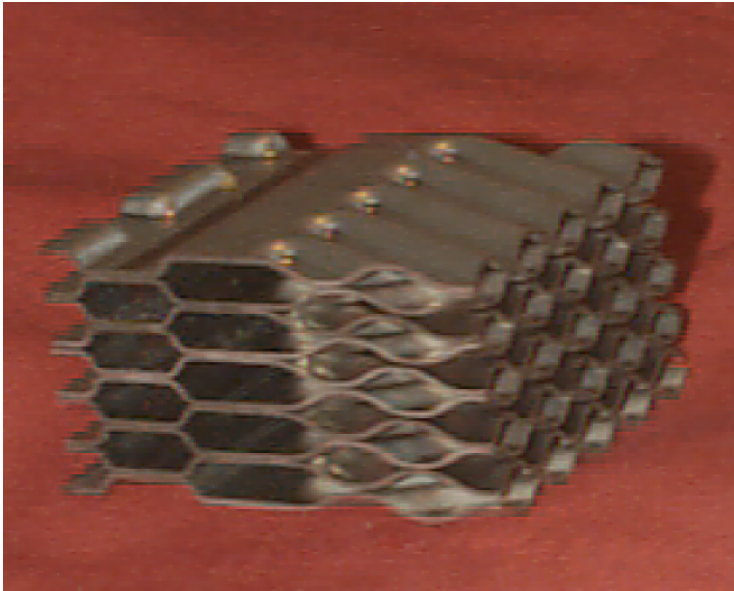
Figure 1.2: A cross-section of a plate heat-exchanger developed and manufactured by Alfa Laval Thermal.

In this application, the influence of four factors on the shape and the quality of the laser weld was investigated. The four factors were *power of laser*, *speed of laser*, *gas flow at nozzle*, and *gas flow at root* (underside) of the welding. The units and settings of low and high levels of these factors are seen in Figure 1.3. The experimenter measured three responses to characterize the shape and the quality of the weld, namely *breakage of weld*, *width of weld*, and *skewness of weld*. These are summarized in Figure 1.4. The aim was to obtain a persistent weld (high value of breakage), of a well-defined width and low skewness.

	Name	Abbr.	Units	Type	Use	Settings
1	**Power**	Po	kW	Quantitative	Controlled	2.15 to 4.15
2	**Speed**	Sp	m/min	Quantitative	Controlled	1.88 to 5
3	**NozzleGas**	No	l/min	Quantitative	Controlled	27 to 36
4	**RootGas**	Ro	l/min	Quantitative	Controlled	27 to 42

Responses		
Name	Abbr.	Units
1 **Breakage**	Br	MPa
2 **Width**	Wi	mm
3 **Skewness**	Sk	-

Figure 1.3: *(left) The four varied factors of General Example 1.*
Figure 1.4: *(right) The three measured responses of General Example 1.*

In the first stage, the investigator carried out eleven experiments. During the data analysis, however, it soon became apparent that it was necessary to upgrade the initial screening design with more experiments. Thus, the experimenter conducted another set of eleven experiments, selected to well supplement the first series. We will provide more details later.

In summary, with a screening design, the experimenter is able to extract a *yes* or *no* answer with regard to the influence of a particular factor. Information is also gained about how to modify the settings of the important factors, to possibly further enhance the result. Screening designs need few experiments in relation to the number of factors.

General Example 2: Optimization

Optimization is used after screening. The objective is (i) to predict the response values for all possible combinations of factors within the experimental region, and (ii) to identify an optimal experimental point. However, when several responses are treated at the same time, it is usually difficult to identify a single experimental point at which the goals for all responses are fulfilled, and therefore the final result often reflects a compromise between partially conflicting goals.

Our illustration of the optimization objective deals with the development of a new truck piston engine, studying the influence on fuel consumption of three factors, *air mass used in combustion*, *exhaust gas re-circulation*, and *timing of needle lift*. The settings of these factors are shown in Figure 1.5. Besides monitoring the *fuel consumption*, the investigator measured the levels of *NOx* and *Soot* in the exhaust gases. These responses are summarized in Figure 1.6. The goal was to minimize fuel consumption while at the same time not exceeding certain stipulated limits of NOx and Soot. The relationships between the three factors and the three responses were investigated with a standard 17 run optimization design. We will provide more details in Chapter 15.

■ Factors					
	Name	Abbr.	Units	Type	Use	Settings
1	Air	Air	kg/h	Quantitative	Controlled	240 to 284
2	EGR%	EGR	%	Quantitative	Controlled	6 to 12
3	NeedleLift	NL	°BTDC	Quantitative	Controlled	-5.78 to 0

■ Responses			
	Name	Abbr.	Units
1	Fuel	Fu	mg/st
2	NOx	NO	mg/s
3	Soot	So	mg/s

Figure 1.5: *(left) The three varied factors of General Example 2.*
Figure 1.6: *(right) The three measured responses of General Example 2.*

In summary, with an optimization design the experimenter is able to extract detailed information regarding how the factors combine to influence the responses. Optimization designs require many experiments in relation to the number of investigated factors.

General Example 3: Robustness testing

The third objective is robustness testing, and it is applied as the last test just before the release of a product or a method. When performing a robustness test of a method – as in the example cited below – the objective is (i) to ascertain that the method is robust to small fluctuations in the factor levels, and, if non-robustness is detected, (ii) to understand how to alter the bounds of the factors so that robustness may still be claimed.

To portray a typical robustness test of an analysis method, we have selected an application taken from the pharmaceutical industry, which deals with a high-performance liquid chromatography (HPLC) system. Five factors, of which four were quantitative and one qualitative, were examined. These factors were *amount of acetonitrile* in the mobile phase, *pH*, *temperature*, *amount of the OSA counterion* in the mobile phase, and *type of stationary phase* (column). These five factors, summarized in Figure 1.7, were investigated using a design of 12 experiments. To describe the chromatographic properties of the HPLC-system, three responses were recorded, that is, the *capacity factor k_1* of analyte 1, the *capacity factor k_2* of analyte 2, and the *resolution Res1* between these two analytes (Figure 1.8).

	Name	Abbr.	Units	Type	Use	Settings
1	AcN	Ac	%	Quantitative	Controlled	25 to 27
2	pH	pH		Quantitative	Controlled	3.8 to 4.2
3	Temp	Te	°C	Quantitative	Controlled	18 to 25
4	OSA	OS	mM	Quantitative	Controlled	0.09 to 0.11
5	Column	Co		Qualitative	Controlled	ColA,ColB

	Name	Abbr.	Units
1	k1	k1	
2	k2	k2	
3	Res1	Re1	

Figure 1.7: *(left) The five investigated factors of General Example 3.*
Figure 1.8: *(right) The three registered responses of General Example 3.*

In HPLC, capacity factors measure the retention of compounds, and resolution the separation between compounds. In the present case, the resolution response was the main interest and required to be robust. More information regarding this example will be given at a later stage (Chapter 17).

In summary, with a robustness testing design, it is possible to determine the sensitivity of the responses to small changes in the factors. Where such minor changes in the factor levels have little effect on the response values, the analytical system is determined to be robust.

The CakeMix application

We will now concentrate on the CakeMix application, which is helpful in illustrating the key elements of DOE. This is an industrial pilot plant application in which the goal was to map a process producing a cake mix to be sold in a box, for instance, at a supermarket or shopping mall. On the box there will be instructions on how to use the cake mix, and these will include recommendations regarding baking temperature and time.

There are many parameters which might affect the production of a cake mix, but in this particular investigation we will only be concerned with the recipe. The experimental objective was screening, to determine the impact of three cake mix ingredients on the taste of the resulting cake. The first varied factor (ingredient) was *Flour*, the second S*hortening (fat)*, and the third *Eggpowder*. In reality, the investigated cake mix contained other ingredients, like sugar and milk, but to keep things simple only three ingredients were varied.

Firstly, the standard operating condition, the center-point, for the three factors was defined, and to do this a recommended cake mix composition was used. The chosen center-point corresponded to 300g Flour, 75g Shortening, and 75g Eggpowder. Secondly, the low and the high levels of each factor were specified in relation to the center-point. It was decided to vary Flour between 200 and 400g, Shortening between 50 and 100g, and Eggpowder between 50 and 100g. Thirdly, a standard experimental plan with eleven experiments was

created. This experimental design is shown in Figure 1.9, and in this table each row corresponds to one cake.

Cake Mix Experimental Plan				
Cake No	**Flour**	**Shortening**	**Egg Powder**	**Taste**
1	200	50	50	3.52
2	400	50	50	3.66
3	200	100	50	4.74
4	400	100	50	5.20
5	200	50	100	5.38
6	400	50	100	5.90
7	200	100	100	4.36
8	400	100	100	4.86
9	300	75	75	4.73
10	300	75	75	4.61
11	300	75	75	4.68

Factors		
	Levels (Low/High)	**Standard condition**
Flour	200 g / 400 g	300 g
Shortening	50 g / 100 g	75 g
Egg powder	50 g / 100 g	75 g
Response: Taste of the cake, obtained by averaging the judgment of a sensory panel.		

Figure 1.9: The experimental plan of the CakeMix application.

For each one of the eleven cakes, a sensory panel was used to determine how the cake tasted. The response value used was the average judgment of the members of the sensory panel. A high value corresponds to a good-tasting cake, and it was desired to get as high value as possible. Another interesting feature to observe is the repeated use of the standard cake mix composition in rows 9-11. Such repeated testing of the standard condition is very useful for determining the size of the experimental variation, known as the replicate error.

Apart from listing all the experiments of the design as a table, it is also instructive to make a graphical presentation of the design. In the CakeMix application, a cube is a good tool to visualize the design and thus better understand its geometry. This is shown in Figure 1.10.
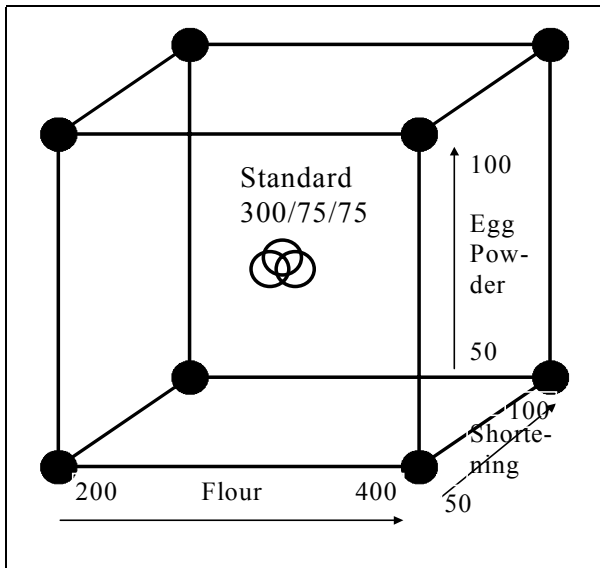
Figure 1.10: *A geometrical representation of the CakeMix experimental protocol.*

After the completion of an experimental plan, one must analyze the data to find out which factors influence the responses. Usually, this is done by fitting a polynomial model to the data. In the CakeMix application, the performed experimental design supports the model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \varepsilon,$$

where y is the response, x's the three ingredients, β_0 the constant term, β's the model parameters, and ε the residual response variation not explained by the model. The model concept, the philosophy of modelling, and model adequacy are further discussed in Chapter 3.

The aim of the data analysis is to estimate numerical values of the model parameters, the so called regression coefficients, and these values will indicate how the three factors influence the response. Such regression coefficients are easy to overview when plotted in a bar chart, and the results for the cake mix data are displayed in Figure 1.11. We see that the strongest term is the two-factor interaction between Shortening and Eggpowder.

Normally, one uses a regression coefficient plot to *detect* strong interactions, but response contour plots to *interpret* their meaning. The response contour plot displayed in Figure 1.12 shows how Taste varies as a function of Shortening and Eggpowder, while keeping the amount of Flour fixed at its high level. Apparently, to obtain a cake with as high "taste" as possible, we should stay in the upper left-hand corner, i.e., use much Flour, much Eggpowder and little Shortening.
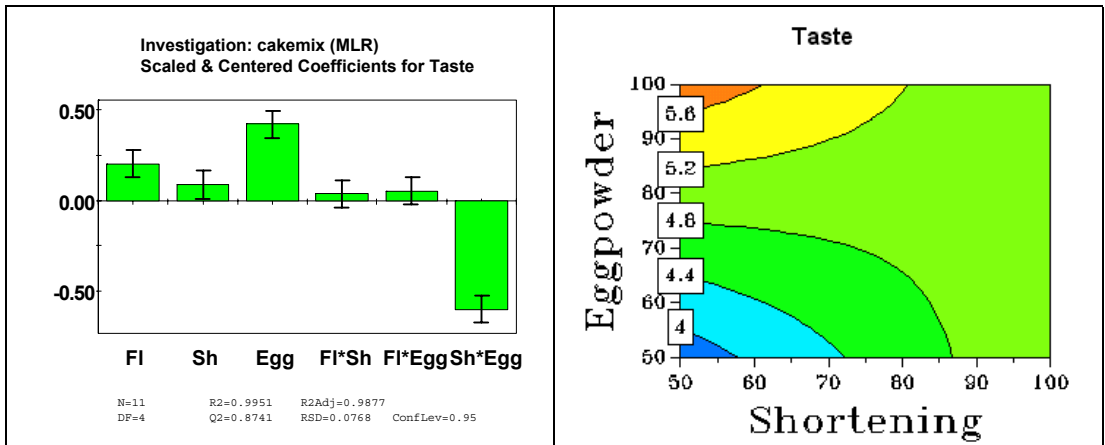
Figure 1.11: (left) Regression coefficient plot of CakeMix regression model.
Figure 1.12: (right) Response contour plot of taste as a function of Eggpowder and Shortening.

The type of response contour plot displayed in Figure 1.12 is useful for decision making – it suggests what to do next, that is, where to continue experimentally. Thus, we have converted the experimental data into an informative map with quantitative information about the modelled system. This is actually the essence of DOE, to plan informative experiments, to analyze the resulting data to get a good model, and from the model create meaningful maps of the system.

Examples of statistical designs

As we have seen, the DOE concept may be viewed as a framework for experimental planning. We shall here briefly overview a few basic designs of this framework, which are used to deal with the three major experimental objectives, and point out their common features and differences. Figure 1.13 provides a summary of the designs discussed.

The first row of Figure 1.13 shows complete, or full, factorial designs for the investigation of two and three factors. These are screening designs, and are called *full* because all possible corners are investigated. The snowflake in the interior part depicts replicated center-point experiments carried out to investigate the experimental error. Usually, between 3-5 replicates are made. The second row in the figure also shows a screening design, but one in which only a fraction of all possible corners have to be carried out. It belongs to the *fractional factorial* design family, and this family is extensively deployed in screening. Fractional factorial designs are also used a lot for robustness testing. The last row of Figure 1.13 displays designs originating from the *composite* design family, which are used for optimization. These are called composite designs because they consist of the building blocks, corner (factorial) experiments, replicated center-point experiments, and axial experiments, the latter of which are denoted with open circles.
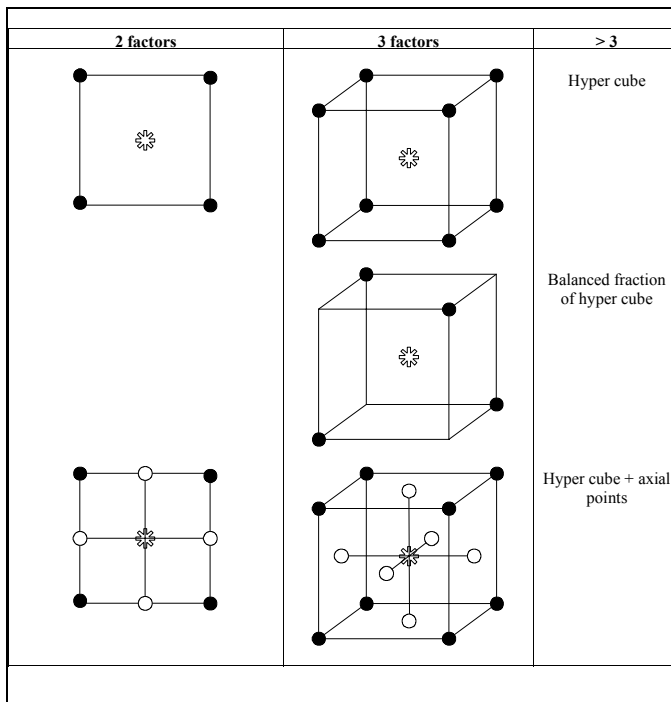
Figure 1.13: Examples of full factorial, fractional factorial, and composite designs used in DOE.

Quiz

Please complete the following statements:

In a screening investigation, the idea is to investigate many factors and their influences on the responses. This is done by using comparatively ….. experiments in relation to the number of varied factors. (many/few)

In an optimization study, one wants to obtain detailed information about how a few factors combine in regulating the responses. This is accomplished by making comparatively ….. experiments in relation to the number of factors. (many/few)

In robustness testing of, for instance, an analytical method, the aim is to explore how sensitive the responses are to small changes in the factor settings. Ideally, a robustness test should show that the responses are *not* sensitive to small fluctuations in the factors, that is, the results are the same for all experiments. Since the expected result is similarity for all runs, a robustness testing design may well be done with …. experiments per varied factor. (very few/very many)

DOE consists of a few well-defined steps. First the experimenter has to select the … and define an …. which can be used to solve the problem. Then the selected experiments have to be carried out and the resulting data analyzed. The data analysis will give a model, which may be interpreted. To understand which factors are most important and to examine whether there are interactions, it is instructive to look at a …. plot. To even better understand the modelled system, one may convert the model information into a map, a … plot, which

explicitly shows where to continue experimentally. (experimental objective/experimental plan/coefficient/response contour)

Benefits of DOE

The great advantage of using DOE is that it provides an *organized* approach, with which it is possible to address both simple and tricky experimental problems. The experimenter is encouraged to select an appropriate experimental objective, and is then guided to devise and perform a *set* of experiments, which is adequate for the selected objective. Although the experimenter may feel some frustration about having to perform a series of experiments, experience shows that DOE requires fewer experiments than any other approach. Since these few experiments belong to an experimental plan, they are mutually connected and thereby linked in a logical and theoretically favorable manner. Thus, by means of DOE, one obtains more useful and more precise information about the studied system, because the joint influence of all factors is assessed. After checking the model adequacy, the importance of the factors is evaluated in terms of a plot of regression coefficients, and interpreted in a response contour plot. The latter type of plot constitutes a map of the system, with a familiar geometrical interpretation, and with which it is easy to decide what the next experimental step ought to be.

Summary

Design of experiments is useful in the laboratory, the pilot plant and full-scale production, and is used for any experimental objective, including screening, optimization, and robustness testing. We have introduced three general examples - the laser welding case, the truck engine study, and the HPLC robustness problem - which will be used to illustrate these three objectives. In addition, the CakeMix application was outlined for the purpose of overviewing some of the key elements involved in DOE. By conducting an informative set of eleven experiments, it was possible to create a meaningful response contour plot, showing how to modify the cake mix recipe to achieve even better tasting cakes. Finally, this chapter ended with a discussion of the main benefits of DOE.

2 Introduction (Level 2)

Objective

The aim of this chapter is two-fold. Initially, we will focus on three critical problems, which are difficult to cope with using the COST-approach, but which are well handled with designed experiments. This problem discussion will automatically lead to the concept of variability, which comprises the second main topic of this chapter. Variability is always present in experimental work, in terms both of the wanted systematic part - caused by changing important factors - and of the unwanted unsystematic noise. It is important to plan the experiments so as to allow estimation of the size of systematic and unsystematic variability. We will discuss the consequences of variability on the planning of the experimental design. The CakeMix application will be used to illustrate how variability is taken into account and interpreted.

Three critical problems

There are three critical problems which DOE handles more efficiently than the COST-approach. The first problem concerns the understanding of a system or a process influenced by many factors. In general, such systems are poorly studied by changing one factor at a time (COST), because interactions between factors cannot be estimated. The DOE-approach, however, enables the estimation of such interactions. Secondly, systematic and unsystematic variability, the former of which is called *effects* and latter of which is called *noise*, are difficult to estimate and consider in the computations without a designed series of experiments. This second problem will be discussed later. Finally, the third critical problem is that reliable maps of the investigated system are hard to produce without a proper DOE-foundation. It is very useful to inspect a reliable response contour plot of the investigated system to comprehend its behaviour. Unfortunately, such a contour plot may be misleading unless it is based on a set of designed experiments. For a response contour plot to be valid and meaningful, it is essential that the experiments have been positioned to well cover the domain of the contour plot. This is usually *not* the case with the COST-approach.

Variability

We will now discuss the concept of variability. Consider Figure 2.1 in which the upper graph displays the yield of a product measured ten times under identical experimental conditions. Apparently, these data vary, despite being obtained under identical conditions. The reason for this is that every measurement and every experiment is influenced by noise. This happens in the laboratory, in the pilot-plant, and in the full-scale production. It is clear that each experimenter must know the size of the experimental noise in order to draw correct conclusions. Indeed,

experimental designs are constructed in such a way that they permit a proper estimation of such noise.

Ten measurements of yield, under identical conditions

Ten measurements of yield as a function of time
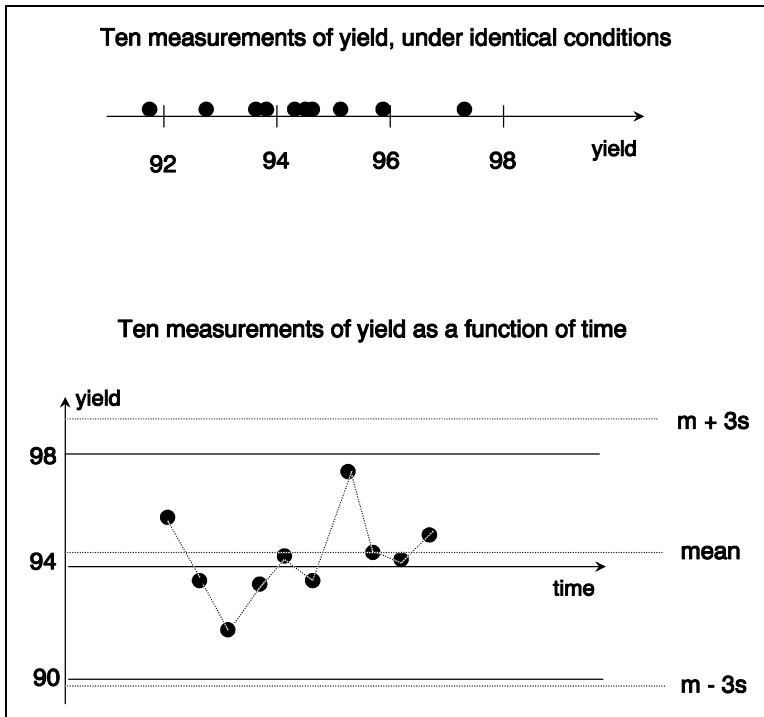
Figure 2.1: *The yield of a product measured ten times under identical experimental conditions.*

Moreover, since the ten measurements were sampled over time, one may use a time series plot showing the variation over time. Such a plot is displayed in the lower graph of Figure 2.1, together with some "control limits" indicating acceptable variation. We see that the data vary in a limited interval and with a tendency for grouping around a central value. The *size* of this interval, usually measured by the *standard deviation*, and the *location* of this central value, usually estimated by the *average*, may be used to characterize the properties of the variability. Once these quantities have been determined they may be used to monitor the behaviour of the system or the process. Under stable conditions, it can be expected that every process and system varies around its average, and stays within the specified control limits.

Reacting to noise

Let us consider the foregoing experimental system from another angle. It was decided to carry out two new experiments and change the settings of one factor, say, temperature. In the first experiment, the temperature was set to 35°C and in the second to 40°C. When the process was operated at 35°C the yield obtained was slightly below 93%, whereas at 40°C the yield became closer to 96%. This is shown in Figure 2.2. Is there any real difference between these two yields, i.e., does temperature have an effect on yield? An experimenter ignorant of the experimental variability of this system would perhaps conclude: Yes, the 5°C temperature increase induces an approximate 3% change in the yield. However, an experimenter who compares the 3% change in yield against the existing experimental variability, would not arrive

at this conclusion. For him it would be obvious that the observed 3% change lies completely within the boundaries established by the 10 replicated experiments. In this case, one cannot be sure about the size of the effect of the temperature. However, it is likely that the temperature *has* an effect on the yield; the question is how large. Thus, what we would like to accomplish is a partition of the 3% change into two components, the *real effect* of the temperature and the *noise*. As will soon be seen, this is best accomplished with DOE.
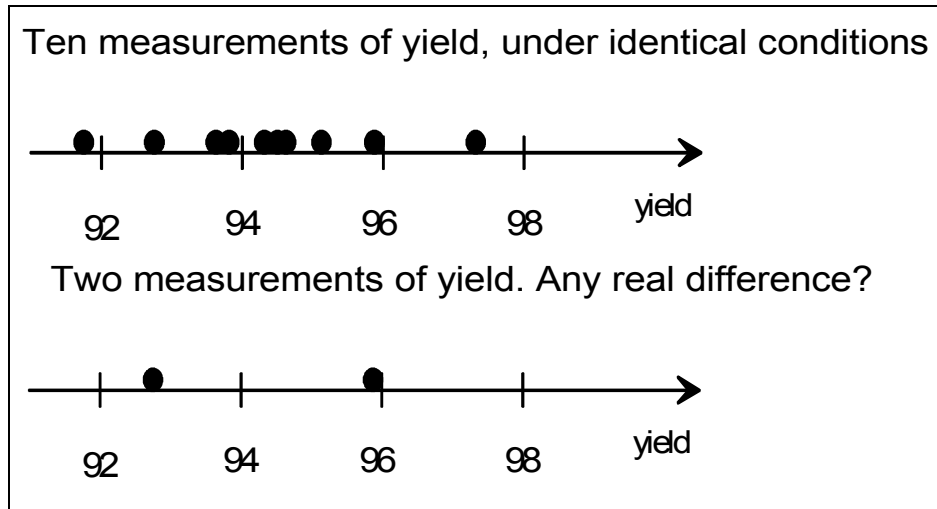
Figure 2.2: An illustration of the effect on yield obtained by changing the settings of one critical factor.

Focusing on effects

Using COST to investigate the "effect" of a factor often leads to a reaction to noise. Another problem with this approach is the number of experiments required, which is illustrated in Figure 2.3. In the left-hand part, five experiments are laid out in a COST fashion to explore the relationship between one factor and one response. This is an informationally inefficient distribution of the experiments. With a better spread, as few as five to seven runs are sufficient for investigating the effects of *two* factors. This is displayed in the center part of the figure. When arrayed in a similar manner, as few as nine to eleven experiments are sufficient for investigating *three* factors, as seen in the right-hand part of the figure. Such square and cubic arrangements of experiments are informationally optimal and arise when DOE is used.
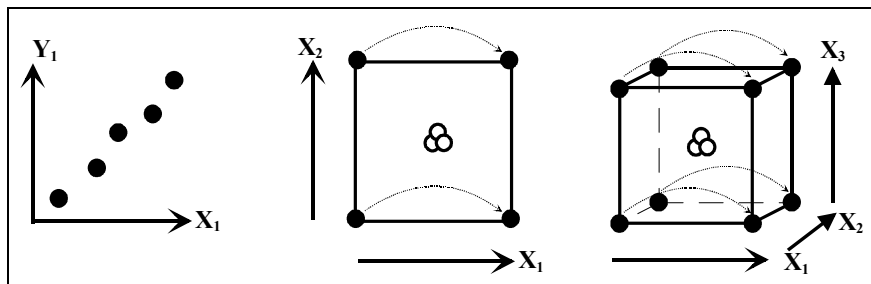
Figure 2.3: The averaging procedure in DOE leading to more precise effect estimates.

The squared arrangement of four experiments has the advantage that we can make *two* assessments of the effect of each factor. Consider the factor x_1. It is possible to estimate the effect of this factor both at the low level of x_2 *and* at the high level of x_2, that is, to study the change in the response when moving along the bottom and top edges of the square. This is schematically illustrated with the two curved, dotted arrows. Each one of these two arrows provides an estimate of the effect of x_1. These estimates are then averaged. The formation of this average implies that the estimated effect of x_1 is sharpened. Analogously, in the three-factor case, the effect of x_1 is computed as the average of four assessments, obtained by moving along the four curved arrows of the cube, and hence the effect of x_1 is well estimated.

Furthermore, with DOE it is possible to obtain an estimate of the noise. This is accomplished by considering the part of the variation which the mathematical model leaves unexplained, the so called residuals. In summary, with DOE it is possible not only to sharpen the estimate of the real effect, thanks to averaging, but also to estimate the size of the noise, e.g. the standard deviation of the residuals. This leads to a focusing on the *real effects* of factors, and not on some coincidental noise effect. In addition, DOE always needs fewer experiments than COST.

Illustration: CakeMix

Given that it is possible to estimate both *real effects* of factors and experimental *noise*, one may wonder *how do we use such estimates in our daily work?* We will exemplify this using the CakeMix example. Figure 2.4 shows the regression coefficients of the interaction model and their confidence intervals. The first three coefficients, also called linear terms, reveal the *real effects* of the three ingredients. The last three coefficients, also called interaction terms, show if there are interactions among the factors. The uncertainty of these coefficients is given by the confidence intervals, and the size of these depends on the size the *noise. Hence, real effects are given by the coefficients, and the noise is accounted for by the confidence intervals.*
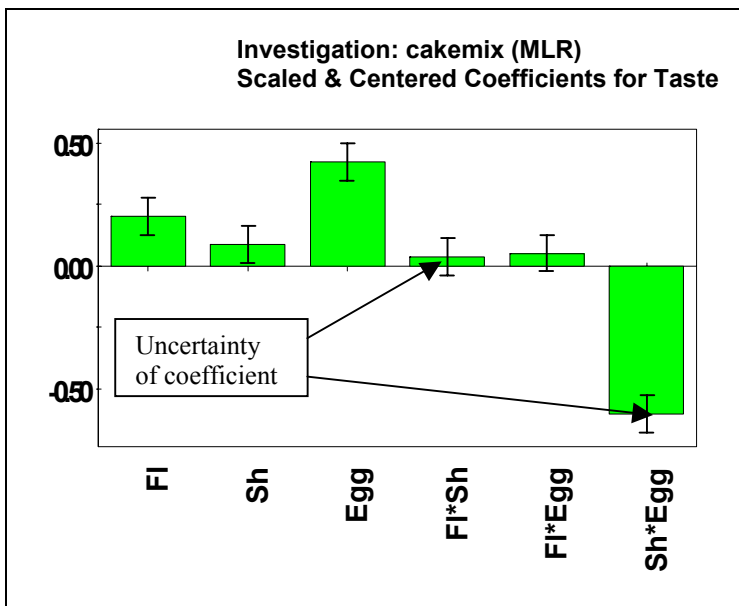
Figure 2.4: *Regression coefficients of the CakeMix interaction model.*

We now turn to the interpretation of the model in Figure 2.4. The linear terms are the easiest ones to interpret, and directly reveal the importance of each factor. We can see that Eggpowder has the strongest impact on the taste. Its coefficient is +0.42, which is interpreted in the following way: When the amount of eggpowder is increased from the standard condition, 75g, to its high level, 100g, and keeping the other factors fixed at their standard condition, the taste of the cake will increase by 0.42. The latter value is expressed in the same unit as the taste response. The second-most influential factor is the amount of flour. When flour is increased from 300g to 400g the taste is modelled to increase by 0.2 unit. The third ingredient, shortening, has comparatively little impact on taste. Interestingly, the most important term in the model is the interaction between shortening and egg. This term will be interpreted later.

In summary, we have here exemplified the partition of *observed effects* into *real effects* and *noise*, and shown how these complement each other in the model evaluation.

Consequence of variability

An important consequence of variability is that it matters a lot *where* the experiments are performed. Consider the top graph of Figure 2.5. In this graph we have plotted an arbitrary response, y, against an arbitrary factor, x. Two experiments have been carried out, and the variability around each point is depicted with error bars. With only two experiments it is possible to calculate a simple mathematical model, a line. When these two experiments are very close to each other, that is, the investigation range of the factor is small, the slope of the line will be poorly determined. In theory, the slope of this line may vary considerably, as indicated in the plot. This phenomenon arises because the model is unstable, that is, it is based on ill-positioned experiments.

However, when these two experiments are far away from each other (upper right-hand graph of Figure 2.5), the slope of the line is well determined. This is because the investigation range of the factor is considerably larger than the experimental variability, and hence there is a strong enough "signal" for the factor to be modelled. Furthermore, with a quantitative factor, it is also favourable to put in an extra point in between, a so called center-point, to make sure that our model is OK. This is shown in the bottom left-hand graph of Figure 2.5. With this third level of the factor, it is possible to determine whether there is a linear or non-linear relationship prevailing between the factor and the response.

Two points, experiments, close to each other make the slope (of the line) poorly determined.	Two points far away from each other make the slope well determined.

And if a center-point is put in between it is possible to explore whether our model is OK. Should it be linear or non-linear?
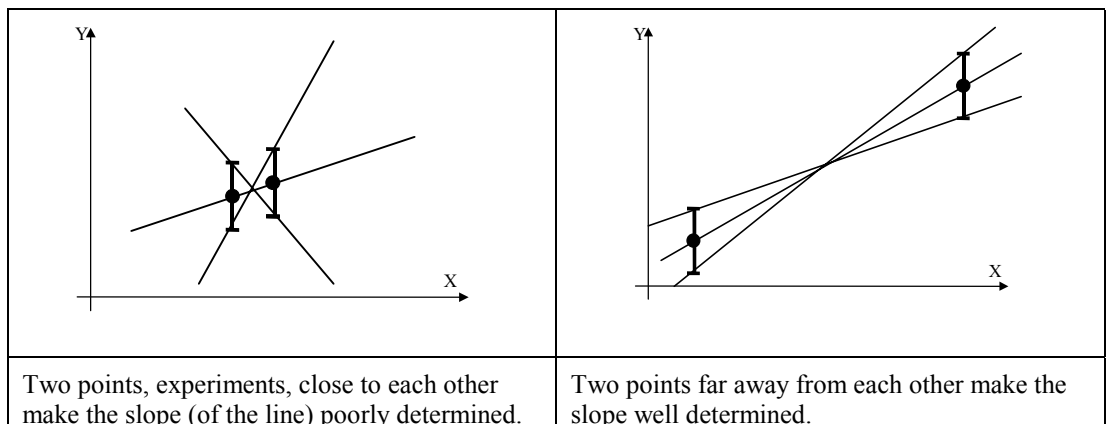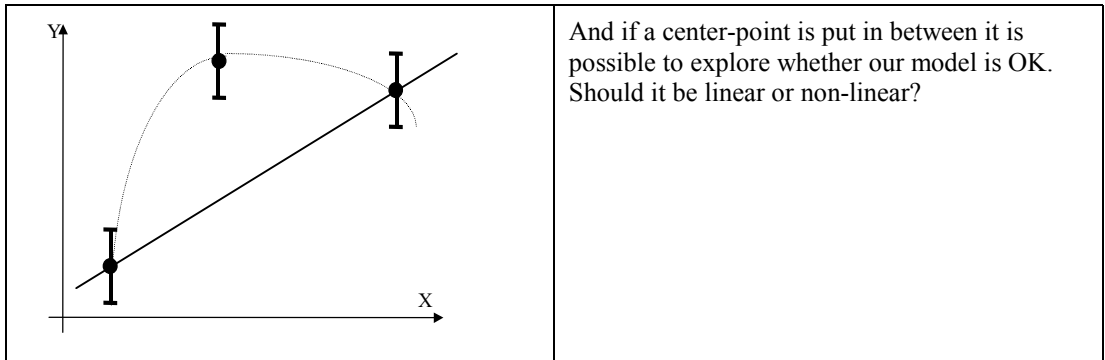
Figure 2.5: *A consequence of variability is that it matters where the experiments are performed.*

Quiz

Please answer the following questions:

Which are the three critical problems that DOE addresses better than the COST-approach?

What is variability?

Is it important to consider variability?

Is it possible to separate observed effects into real effects and noise with COST? With DOE?

Why is it important that the range of a factor is made sufficiently large?

Why is it useful to also use center-point experiments in DOE?

Summary

Design of experiments, DOE, efficiently deals with three critical problems, where the COST-approach is inefficient or unsuccessful. These three problems concern (i) how to monitor a system simultaneously influenced by many factors, (ii) how to separate observed effects into real effects and noise, and (iii) how to produce reliable maps of an explored system.

Furthermore, in this chapter we have introduced the concept of variability. Emphasis has been placed on the importance of well resolving the real effect of a factor and get it "clear" of the noise, to avoid reacting to just noise. This was demonstrated using the CakeMix application. Here, the calculated regression coefficients relate to the real effects, while the confidence intervals inform about the noise.

3 Introduction (Level 3)

Objective

In this chapter, the aim is to describe the concept of mathematical models. We will discuss what such models are and what they look like. A mathematical model is an approximation of reality, which is never 100% perfect. However, when a model is informationally sound, the researcher is equipped with an efficient tool for manipulating the reality in a desired direction. In addition, we shall discuss the relationship between the investigation range of factors and model complexity, and conclude that quadratic polynomial models are sufficiently flexible for most practical cases.

Connecting factors and responses

In DOE, there are two fundamental types of variables, *factors* and *responses* (see Figure 3.1). The responses inform us about properties and general conditions of the studied system or process. Putting it loosely, one may say that they reveal whether the system behaves in a healthy or unhealthy manner. Typical responses might be taste of cake, stability of weld, fuel consumption of truck engine, resolution of analytical peaks in liquid chromatography, and so on. The factors, on the other hand, are our tools for manipulating the system. Since they exert an influence on the system, the nature of which we are trying to map, it is usually possible to force the system towards a region where it becomes even healthier. Typical factors might be amount of flour in cake mix recipe, gas flow at nozzle of welding equipment, exhaust gas re-circulation in truck engine, pH of mobile phase in liquid chromatography, and so on.

Once factors have been chosen and responses measured, it is desirable to get an understanding of the relationships between them, that is, we want to connect the information in the factor changes to the information in the response values. This is conveniently done with a mathematical model, usually a polynomial function. With such a model it is possible to extract clues like: *to maximize the third response factor 1 should be set high and factor 2 be set low; to minimize the fourth response just lower all factors; the first response is not explicable with the factors studied; there is a non-linear relationship between response 2 and factor 3*, and so on. Such insights are invaluable when it comes to specifying further experimental work.
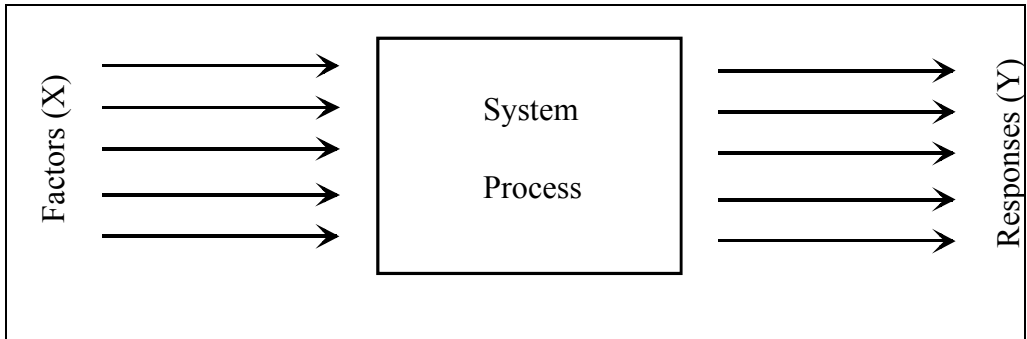
Figure 3.1: The measured responses describe the properties of the investigated system. By changing the most influential factors the features of the system might be altered according to a desired response profile.

The model concept

It is of utmost importance to recognize that a model is an *approximation*, which simplifies the study of the reality. A model will never be 100% perfect, but still be very useful. To understand the model concept, some simple examples are given in Figure 3.2. Consider the electric toy train, something which most children have played with. Such a toy train mimics the properties of a full-sized train, though in a miniature format, and may thus be regarded as a model. Another example is found in geography. For instance, a geographic map of Iceland, with volcanoes and glaciers marked, is a useful and necessary model of reality for tourists. Also, the response contour plot of taste in the CakeMix application is a model, which suggests how to compose a cake mix to get a tasty cake. These three examples well document what is meant with a model. Models are not reality, but approximate representations of some important aspects of reality. Provided that a model is sound – there are tools to test this – it constitutes an excellent tool for understanding important mechanisms of the reality, and for manipulating parts of the reality according to a desired outcome.
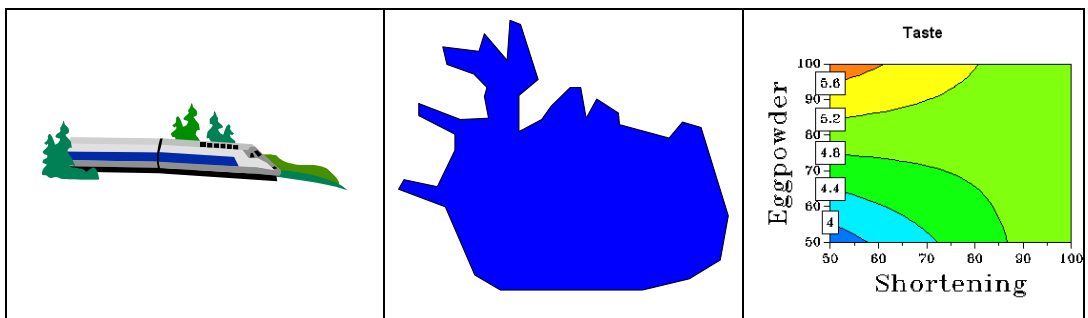
Figure 3.2: Three examples of models, a toy train, a map of Iceland, and a response contour plot of taste.

Empirical, semi-empirical and theoretical models

In this course book, we will work with *mathematical* models. Certain classes of mathematical models are discernible, i.e., *empirical models*, *semi-empirical models*, and

theoretical models (see Figure 3.3). A theoretical model, also called a *hard* model, is usually derived from a well-established and accepted theory within a field. Consider, for example, the Schrödinger equation, which is a typical representative of this class of models. Theoretical models are often regarded as fundamental "laws" of natural science even though the label "models" would be more appropriate in experimental disciplines. In most cases, however, the mechanism of a system or a process is usually not understood well enough, or maybe too complicated, to permit an exact model to be postulated from theory. In such circumstances, an empirical model based on experiments, also called a *soft* model, describing how factors influence the responses in a local interval, might be a valuable alternative.

Further, in DOE, the researcher often has *some* prior knowledge that *certain* mathematical operations might be beneficial in the model building process. For instance, we all know that it is not particularly meaningful to work with the concentration of the hydrogen ion in water solution. Rather, it is more tractable to work with the negative logarithm of this concentration, a quantity normally known as pH. Another example might be that the experimenter knows that it is not the factors A and B that are interesting per se, but that the ratio B/A is what counts. When we factor this kind of prior knowledge into an empirical investigation, we are conducting *partial* empirical modelling, and consequently the prefix semi- is often added. Because of this, it is often stated that DOE involves semi-empirical modelling of the relationships between factors and responses.

Empirical	Semi-empirical	Fundamental
$y = a + bx + \varepsilon$	$y = a + b\log x + \varepsilon$	$H\psi = E\psi$

Figure 3.3: An overview of empirical, semi-empirical and fundamental mathematical models.

Semi-empirical modelling – Taylor series expansions

Taylor series expansions are often used in semi-empirical modelling. Consider Figure 3.4 in which the "true" relationship between one response and one factor, the target function $y = f(x)$, is plotted. We will try to find an alternative function, $y = P(x)$, which can be used to approximate the true function, $y = f(x)$. In a limited factor interval, Δx, any continuous and differentiable function, $f(x)$, can be arbitrarily well approximated by a polynomial, $P(x)$, a Taylor series of the form: $y = P(x) = b_0 + b_1 x + b_2 x^2 + ... + b_p x^p + e$, where e represents a residual term. In this polynomial function, the degree, p, gives the complexity of the equation.
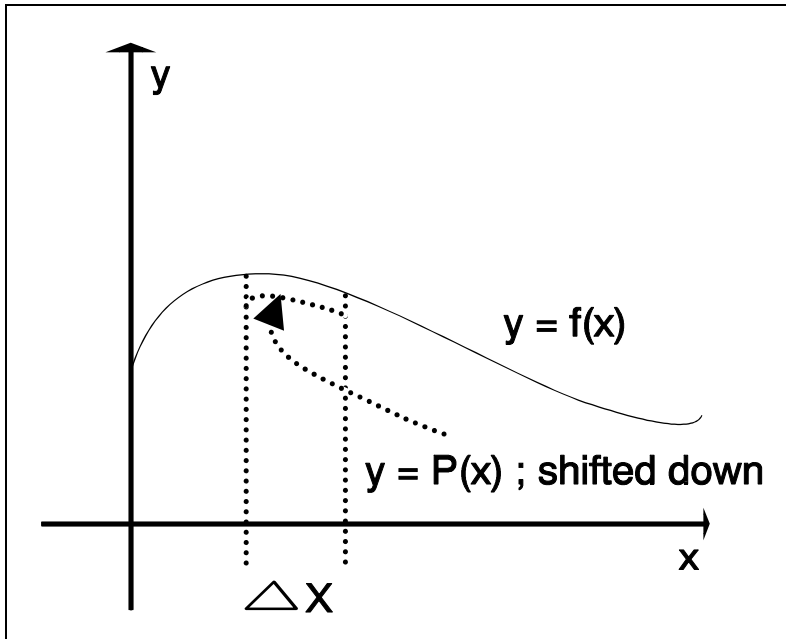
Figure 3.4: *The "true" relationship, y = f(x), between a response, y, and a factor, x, approximated by an alternative function, y = P(x), in a limited factor interval Δx.*

Because there exists an important relationship between the factor interval, Δx, and the degree of the polynomial model, p, it is worthwhile to carry out a closer inspection of Figure 3.4. Let us assume that the degree p has been fixed. Then, for this degree, p, the approximation of y = f(x) by y = P(x) gets better the smaller the factor interval, Δx, becomes. Analogously, for a fixed size of the factor interval, Δx, the approximation is better the higher the degree, p, is. Hence, what is important to always bear in mind in modelling is this trade-off between model complexity and investigation ranges of factors. Of course, this kind of reasoning can be generalized to functions of many factors and many responses.

Conceptual basis of semi-empirical modelling

A semi-empirical model is a *local* model, which describes in detail the situation within the investigated interval. This is a desirable feature. We do not want a *global* model only providing superficial knowledge of larger aspects of reality. A small example will explain why. Imagine a tourist wanting to find the exact position of the Eiffel Tower in Paris. Surely, this person would not look at a globe of our planet, because it represents too global a model, with little or no useful information of where to find the Eiffel Tower. Not even a map of France is local enough, but a map of the central parts of Paris would do the job well. Such a town map of Paris represents a local model of a small aspect of reality, and provides detailed information of where to find the great tourist attraction.

Models of complicated systems or processes in science and technology function in the same way. Local models pertaining to narrowly defined investigation regions provide more detail than models covering large regions of seemingly endless character. This is one reason why it is important to carefully specify the investigation range of each factor. Interestingly, our long experience in technology, chemistry, biology, medicine, and so on, shows that within

such a restricted experimental area nature is fairly smooth and not extensively rugged (see Figure 3.5). This means that when several factors are explored, a smooth, waving response surface may typically be encountered. Such a smooth response surface can be well approximated by a simple polynomial model, usually of quadratic degree.
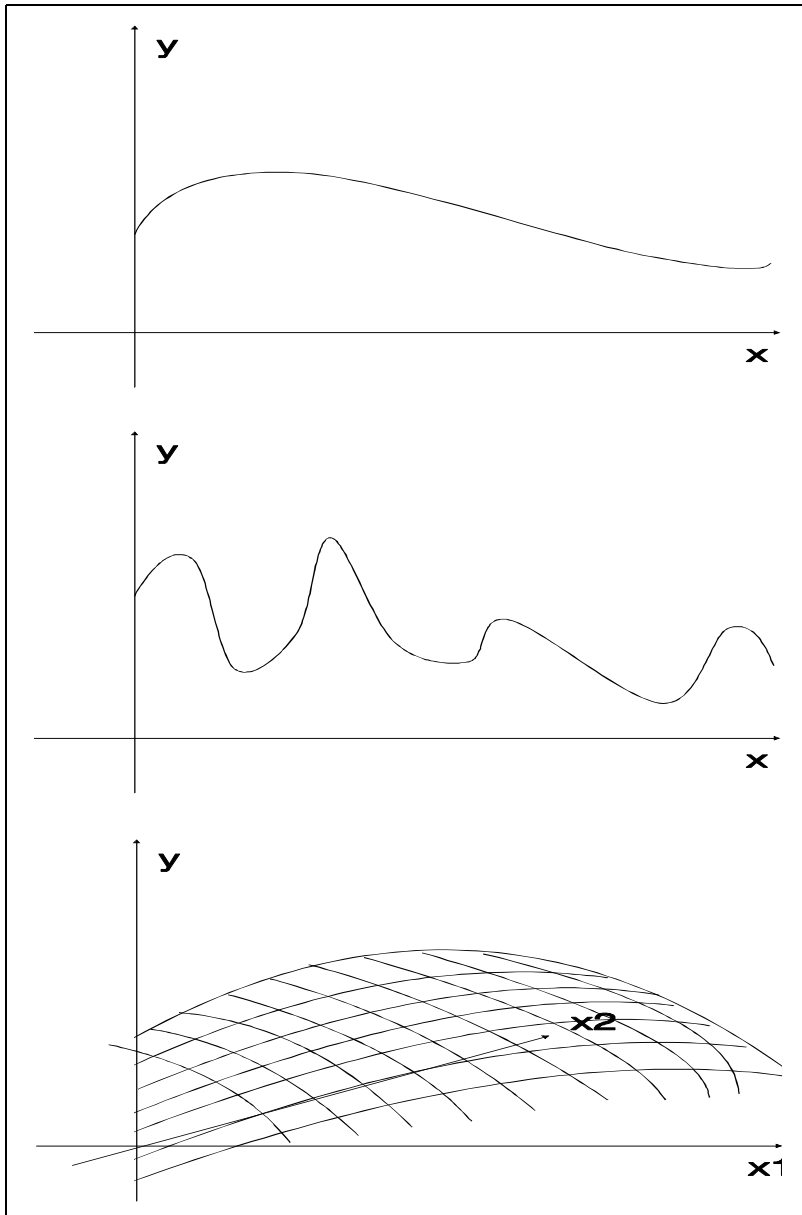
Figure 3.5: Our long experience (technology, chemistry, biology, medicine,…) shows that nature is fairly smooth and not rugged. With several factors a smooth response surface is usually applicable. Such a smooth surface can be well approximated by a quadratic polynomial model.

Quiz

Please answer the following questions:

Why is it necessary to obtain a model linking factors and responses together?

What is a model?

Is a model always a 100% perfect representation of reality?

What is a theoretical model?

What is a semi-empirical model?

What can you say about the trade-off between model complexity and size of factor ranges?

Why is a local model preferable to a global model?

Summary

In order to understand how factors and responses relate to each other, and to reveal which factors are influential for which responses, it is favorable to calculate a polynomial model. It is important to recall that such a model is a simplification of some small aspects of reality, and that it will never be 100% perfect. However, with a sufficiently good model, we have an efficient tool for manipulating a small part of reality in a desired direction. In DOE, we work with semi-empirical modelling, and the models that are calculated have a local character, because they are applicable in a confined experimental region. Because we are investigating reality in a local interval, where nature is often smooth, simple low-order polynomials up to quadratic degree are usually sufficient.

4 Problem formulation (Level 1)

Objective

The problem formulation step is of crucial importance in DOE, because it deals with the specification of a number of important characteristics critically influencing the experimental work. In the problem formulation, the experimenter must specify the experimental objective, define factors and responses, choose a regression model, and so on, and any mistakes at this step may result in difficulties at later stages. The aim of this chapter is to overview how a correct problem formulation is carried out. A secondary aim is to outline six important stages in the experimental process. These are (i) familiarization, (ii) screening, (iii) finding the optimal region, (iv) optimization, (v) robustness testing, and (vi) mechanistic modelling.

Problem formulation

The problem formulation is of importance in DOE, regardless of whether the application concentrates on screening, optimization, or robustness testing. The objective of carrying out the problem formulation, or problem definition, is to make completely clear, for all involved parties, the intentions underlying an experimental investigation. There are a number of things to discuss and agree about, and it is necessary to consider six points. These six items regard (i) experimental objective, (ii) factors, (iii) responses, (iv) model, (v) design, and (vi) worksheet.

The experimental objective defines what kind of investigation is required. One should ask *why is an experiment done*? And *for what purpose*? And *what is the desired result*? The factors are the variables that are changed to give different results on the measured responses. The fourth point, model, means that one has to specify a polynomial model that corresponds to the chosen experimental objective. Next, an experimental design is created which supports the selected model. Thus, the composition of the design follows from the experimental objective and the polynomial model, but also depends on the shape of the experimental region and the number of experiments.

When a design has been proposed, it is important to sit down and go through the proposed experiments to make sure that all of them look reasonable, can be performed, and have the potential of fulfilling the experimental objective. Subsequently, the experimental worksheet is created. In general, such a worksheet is similar to the actual design. However, in the worksheet, information of practical experimental character is normally appended, such as a proposed random run order, and supplementary experiments already carried out.

In the next seven sections, we will discuss the selection of the experimental objective (item (i)), and towards the end of this chapter describe the other five important points (items (ii)-(vi)) in more detail.

Stages in the experimental process

We will now describe the first step of the problem formulation, the selection of the experimental objective. Basically, the experimental objective may be selected from six DOE-stages: (i) familiarization, (ii) screening, (iii) finding the optimal region, (iv) optimization, (v) robustness testing, and (vi) mechanistic modelling. We will highlight the merits and requirements of these stages. However, we will *not* enter into any details regarding mechanistic modelling, as it rarely occurs in industrial practice. In summary, for each one of these stages/objectives, the experimental design will vary depending on the type of factors, the number of factors, possible constraints (restrictions), and the selected model.

Familiarization

The familiarization stage is used when faced with an entirely new type of application or equipment. Many researchers, who are already familiar with their applications, may skip this step and go directly to the next step, screening. In the familiarization stage, one normally spends only a limited amount of the available resources, say 10% at maximum. With the term *resources* we refer to money, time, personnel, equipment, starting materials, and so on. The basic idea is to obtain a brief feeling for what is experimentally possible, and how one should proceed to exert some kind of influence on the investigated system or process. In other words, it is desirable to find out whether the process is possible at all, and, if it is possible and working properly, roughly how good it is.

Familiarization may be accomplished using chemical and technological insight, intuition, and very simple designs. Such a simple design is usually selected from the factorials family and is predominantly constructed in the two factors believed to be most important. As seen in Figure 4.1, this factorial design may be supplemented with some center-point experiments. Usually, such a design is constructed only to verify that similar results are obtained for the replicated center-points, and that different results are found in the corners of the design.
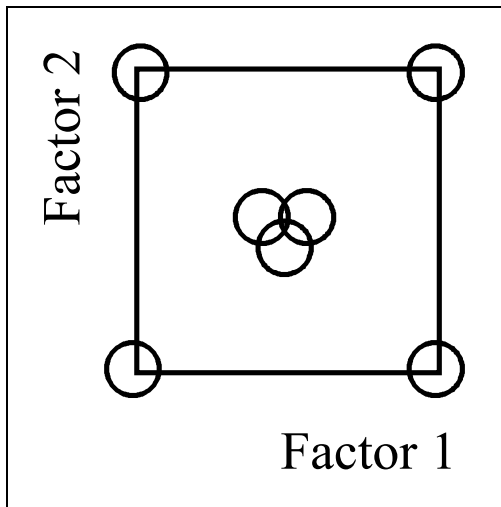
Figure 4.1: A simple factorial design useful in familiarization. Usually, such a design is constructed only to verify that similar results are obtained for the replicated center-points, and that different results are found in the corners of the design.

In addition, another main aim of this stage concerns the development of response measurement techniques. One should always choose responses that are relevant to the goals set up in the problem formulation, and many responses are often necessary. In this stage, we will therefore have to verify that our response measurement techniques work in reality.

Screening

The second experimental stage is screening, which in this course is highlighted as one of the three major experimental objectives (screening, optimization, and robustness testing). With screening one wants to find out a little about *many* factors, that is, which factors are the dominating ones, and what are their optimal ranges? Interestingly, the 80/20 rule, also known as the Pareto principle after the famous Italian economist Vilfredo Pareto, applies well to this stage. In the context of DOE, the Pareto principle states that 80% of the effects on the responses are caused by 20% of the investigated factors. This is illustrated in Figure 4.2, in which approximately 20% of the investigated factors have an effect exceeding the noise level. These strong factors are the ones we want to identify with the screening design. Another way to see this is exemplified in Figure 4.3. Prior to putting the screening design into practice, the examined factors are ascribed the same chance of influencing the responses, whereas after the design has been performed and evaluated, only a few important factors remain. With the screening objective, simple linear or interaction polynomials are sufficient, and the designs employed are primarily of the fractional factorial type.
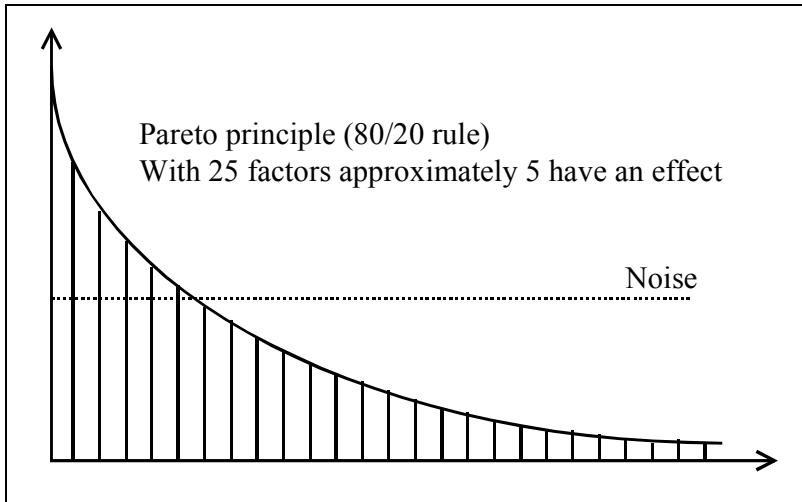
Figure 4.2: The Pareto principle in DOE suggests that 20% of the factors have an effect above the noise level.
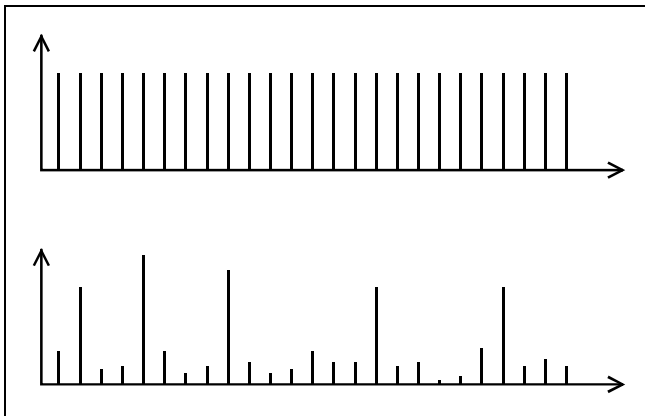
Figure 4.3: Prior to the screening examination all factors are ascribed the same chance of influencing the responses, and after the screening only a few dominating ones remain.

Finding the optimal region

Occasionally, after the screening stage, one has to conclude that the experimental region explored is unlikely to contain the optimum. In such a situation, it is warranted to move the experimental region into a domain more likely to include the desired optimal point. The question which arises, then, is *how do we move the experimental region to an appropriate location*? This is illustrated in Figure 4.4. The tricky part lies in the maneuvering of the experimental region up to the top of the mountain.

In reality, a re-positioning of the experimental region is accomplished by using the outcome of the existing screening design, that is, the polynomial model, which dictates in which direction to move. This is exemplified in Figure 4.5. The interesting direction, which lies perpendicular to the level curves (parallel lines) of the response contour plot, is used for moving according to the principles of *steepest ascent* or *steepest descent*, depending on

whether a maximum or minimum is sought. The example shows a steepest descent application.
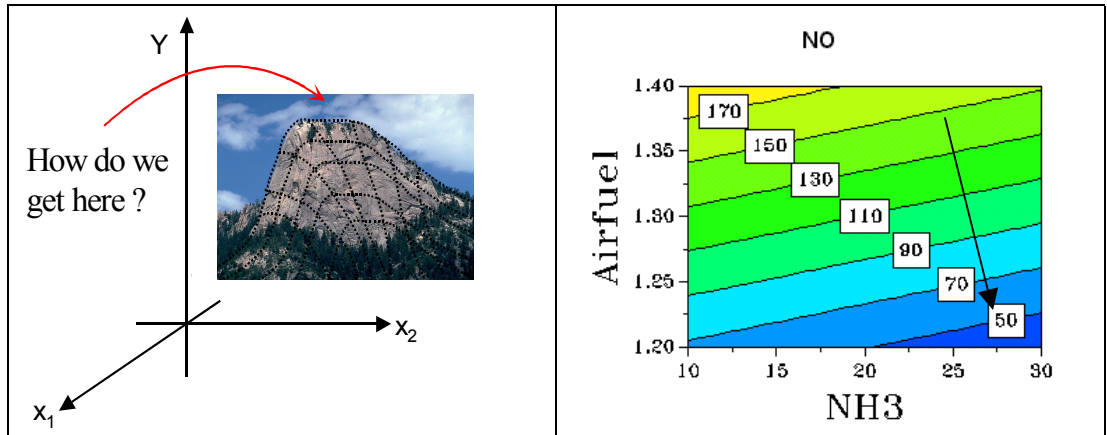
Figure 4.4: (left) After the screening stage, one is often interested in moving the experimental region so that it includes the optimum.
Figure 4.5: (right) To accomplish a re-positioning of the experimental region, one uses the results from the screening design to move in the direction of steepest ascent for maximum or steepest descent for minimum.

We emphasize that these gradient techniques are greatly facilitated by the use of graphical tools, such as response contour plots. The usual result of this stage is that one obtains a new experimental point which is suited for anchoring an optimization design, but sometimes it is found that an entirely new screening design is necessary.

Optimization

The next experimental stage is optimization. Now, our perspective is that the most important factors have been identified, as a result of the screening phase, and that the experimental region is appropriately positioned so that it presumably contains the optimal point. Before we discuss optimization in more detail, we shall contrast screening and optimization. At the screening stage, *many* factors are investigated in *few* runs, but in optimization the reverse is true, because *few* factors are explored in comparatively *many* experiments. In addition, we also have to re-phrase our main question. In screening, we ask *if* a factor is relevant, and the expected answer is either yes or no. In optimization, we ask *how* a factor is important, and expect an answer declaring whether there is a positive or negative relation between that factor and a response. This answer must also reveal the nature of such a relationship, that is, is it linear, quadratic, or maybe even cubic?

Hence, we may say that in optimization the main objective is to extract in-depth information about the *few* dominating factors. This is achieved with designs of the *composite* family, which support quadratic polynomial models by encoding between 3 and 5 levels of each factor. A quadratic model is flexible and may closely approximate the "true" relation between the factors and the responses. A convenient way of overviewing the implication of a fitted model is to display the modelling results in terms of a response surface plot. Figure 4.6 shows an example of a response surface plot. Hence, this approach is known as *response surface modelling*, or RSM for short.
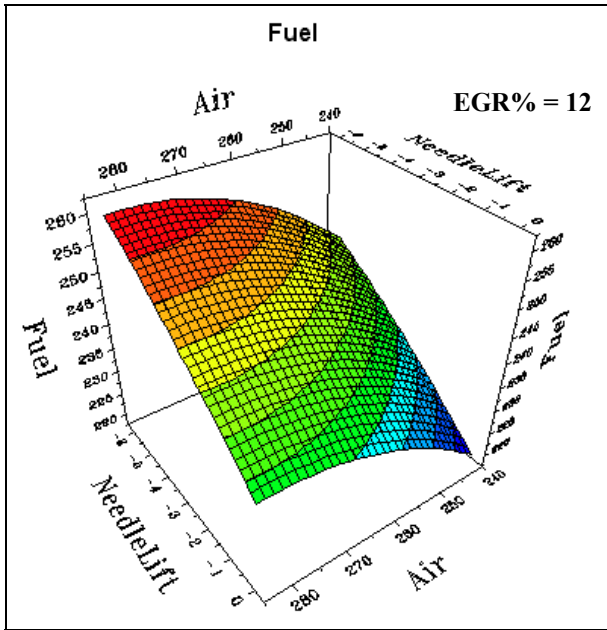
*Figure 4.6: In optimization, the main objective is to extract in-depth information about the **few** dominating factors. A quadratic RSM model is usually used, closely approximating the "true" relation between the factors and the response(s): $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 + ... + \varepsilon$.*

When an RSM model has been fitted it can be used for two primary purposes. The first is the prediction of response values for any factor setting in the experimental region. The second is the identification of the factor setting corresponding to the optimal point, i.e., optimization. With few responses, say 2 or 3, finding the optimum is an easy task. But with more responses, and with sometimes conflicting demands, the identification of an optimal point may be cumbersome, and the final result is often a compromise between various objectives.

Robustness testing

Robustness testing is an important, but often overlooked, stage in the experimental process. It is usually carried out before the release of an almost finished product, or analytical system, as a last test to ensure quality. Accordingly, we highlight robustness testing as one of the three primary experimental objectives. With robustness testing one wants to identify those factors which *might* have an effect on the result and regulate them in such a manner that the outcome is within given specifications. Here, we think of factors that we normally vary in a statistical design, such as, temperature, pressure, gas flow, and so on. Likewise, we want to reveal those factors which in principle have *little* effect, but which may still cause an undesired spread around the ideal result. Typical examples of such factors are ambient temperature, humidity, variability in raw material composition, etc. We want to understand their impact and try to adjust all factors so that the variation in responses is minimized. There are different ways to tackle a robustness testing problem. Usually, a fractional factorial design or Plackett-Burman design is used. This topic will be pursued in more detail in Chapter 17.

Mechanistic modelling

Mechanistic modelling is the last of the stages in the experimental process that we will describe. This stage is attempted when there is a need to establish a theoretical model within a field. One or more semi-empirical models are then utilized in trying to build such a theoretical model. However, it is important to remember that the semi-empirical models are *local* approximations of the underlying fundamental relationships between the factors and the responses. As such, they will never be 100% perfect, which is a source of uncertainty when trying to formulate an exact mechanistic model. In this conversion process, the coefficients of the semi-empirical models are used to get an idea of appropriate derivative terms in the mechanistic model, and it is also crucial that the shape of the former conforms with the shape of the latter, more fundamental relationship. One may use one or many semi-empirical models to attempt to falsify, or maybe prove, a number of competing theoretical models. However, the success of the mechanistic modelling relies heavily on the use of a correct problem formulation, a correct experimental design, and a correct data analysis. The extent to which mechanistic modelling is approached in this way in industry is limited.

Quiz I

Please answer the following questions:

Which six points are important in the problem formulation?

Which are the six stages in the experimental process?

Of these six stages, three are more important as experimental objectives than the others. Which?

What is important regarding the responses at the familiarization stage?

What is the objective in the screening stage?

What kind of polynomial model is sufficient for screening?

How are gradient techniques applied in the search for the optimal experimental region?

What is the main objective in the optimization stage?

What is an RSM model? How can it be interpreted?

What are the main differences between screening and optimization?

When is robustness testing relevant?

Specification of factors

We have now completed the description of the six experimental stages in DOE. The experimenter should select one of these stages as part of the problem formulation. Subsequently, the next step in the problem formulation involves the specification of which factors to change. We will now conduct a general discussion related to the specification of factors.

The factors are the variables which, due to changes in their levels, will exert an influence on the system or the process. In the broadest sense, factors may be divided into *controllable* and *uncontrollable* factors (see Figure 4.7). Controllable factors are the easiest ones to handle and investigate, and the experimenter is usually alerted when such a factor changes. Uncontrollable factors are factors which are hard to regulate, but which may have an impact

on the result. Typical examples of the latter are ambient temperature and humidity. We strongly recommend that such uncontrollable factors be kept track of.

Another way to categorize factors is according to the system of *process factors* and *mixture factors*. Process factors are factors which can be manipulated independently of one another, and which may be investigated using factorial and composite designs. Typical examples are the power of the laser welding equipment, the air mass used in combustion in a truck engine, the pH in the mobile phase of a liquid chromatography system, and so on. Mixture factors are factors which display the amounts of the ingredients, or constituents, of a mixture, and they add to 100%. Hence, mixture factors can *not* be varied independently of one another, and therefore require special designs other than factorial and composite designs.

A third, and perhaps the most common way, of dividing factors is to consider them as either *quantitative* or *qualitative*. A quantitative factor is a factor which may change according to a continuous scale. The three process factors just mentioned are good examples from this category. A qualitative factor is a categorical variable, which can only assume certain discrete values. A typical qualitative factor might be the kind of flour used in making a cake mix, which could, for instance, assume the three levels flour supplier A, flour supplier B, and flour supplier C. In the MODDE software, all these types of factors can be specified.

Controlled

Quantitative (Continuous)

	low	center	high
Temperature	35ºC	40ºC	45ºC
Amount	2g	4g	6g
Speed	200 rpm	250 rpm	300 rpm
pH	5	6	7

Qualitative (Discrete)

Catalyst	Pd	Pt	
Flour supplier	A	B	C
Type of car	Volvo	Saab	Fiat

Uncontrolled

Outside/Inside Temperature

Outside/Inside Moisture

Figure 4.7: In the specification of quantitative factors, it is mandatory to specify a low and a high investigation level. It is recommended to experiment also at a center level, so called center-points, located half-way between the low and high settings. In the specification of qualitative factors, all levels of each factor must be given. It is practical to define between 2 and 5 levels for a qualitative factor. In MODDE, it is possible to explore as many as 10 levels of a qualitative factor, however, this requires many experiments. It is not possible to define true center-point experiments with qualitative factors. It is also favorable in the factor specification to define which factors are uncontrollable. This is done in order to keep track of their values at the time of the experimentation. In the data analysis it may be investigated whether the uncontrollable factors actually have an effect on the responses.

Apart from deciding whether a factor is quantitative, qualitative, or uncontrollable, and so on, the specification of factors also includes defining the investigation range of each factor.

For a quantitative factor it is common practice to define a low and a high level. But, since it is favorable to use replicated experiments in the interior part of a design, defining a center level, located half-way between the low and high levels, is a good practice. This is exemplified in Figure 4.7.

When the low and the high levels of a quantitative factor have been set, they define the investigation *range* of that particular factor. The range of a quantitative factor predominantly depends on three criteria: experimental feasibility, experimental objective, and experimental noise. With experimental feasibility we mean that it is important to make clear which settings are *relevant* for the problem at hand. For instance, when taking a bath, the temperature of the bathing water is of great significance. Theoretically, this temperature may be varied between 0º and 100º C, but it is only relevant to consider the temperature range 35-45ºC.

Also, the experimental objective influences the range of a factor. In screening, one normally maps larger intervals as it is not certain beforehand where the best settings are found. And in optimization, it is possible to narrow the ranges, since one then has a good idea of where to encounter the optimal point. Finally, it is important to consider the experimental error. It is crucial to make the investigation range large enough to allow the effect of each factor to be captured.

For a qualitative factor, which can only assume certain discrete levels, one must specify the exact number of levels for each factor. It is practical to examine a qualitative factor at 2-5 levels, but with many more levels, the number of required experiments increases drastically.

Specification of responses

The next step in the problem formulation is the specification of responses. Note that some thought may have been given to this point during the familiarization stage. In this step, it is important to select responses that are relevant according to the problem formulation. Many responses are often necessary to map well the properties of a product or the performance characteristics of a process. Also, with modern regression analysis tools, it is not a problem to handle many responses at the same time.

There are three types of responses in MODDE, regular, derived and linked. A *regular* response is a standard response measured and fitted in the current investigation. A *derived* response is an artificial response computed as a function of the factors and/or regular or linked responses. A *linked* response is a response that is invoked in the current application, but was defined in another project. The option of linking responses to the ongoing investigation makes it possible to fit separate models to fundamentally different responses and to optimize them together.

When specifying regular responses one must first decide whether a response is of a quantitative or qualitative nature (see Figure 4.8). Examples of the former category are the breakage of a weld, the amount of soot released when running a truck engine, and the resolution of two analytical peaks in liquid chromatography. Quantitative responses are easier to handle than qualitative, because interpretation of regression models is rendered easier.

Sometimes, however, the experimenter is forced to work with qualitative responses. The general advice here is that one should strive to define a qualitative response in as many levels as possible. Let us say that we are developing a product, and that we are unable to measure its quality with a quantitative measurement technique. The only workable option is

a visual quality inspection. Then it is preferable to try to grade the quality in as many levels as is found reasonable. A classification of the product quality according to the five levels worthless/bad/OK/good/excellent is more tractable than a mutilated yes/no type of answer (Figure 4.8). Human quality inspection requires a lot of training, but for instance a well qualified sensory panel can grade the taste and flavor of cheese according to a nine-grade scale.

Quantitative
breakage of weld
soot release when running a truck engine
resolution of two adjacent peaks in liquid chromatography
Qualitative
categorical answers of yes/no type
the cake tasted good/did not taste good
Semi-qualitative:
Product quality was
Worthless = 1
Bad = 2
OK = 3
Good = 4
Excellent = 5

Figure 4.8: *Categorization of responses and some examples.*

Selection of model

The next step in the problem formulation is the selection of an appropriate regression model. This selection is an integral part of the problem formulation. We distinguish between three main types of polynomial models, which are frequently used. These are *linear*, *interaction*, and *quadratic* polynomial models (Figure 4.9).

Linear
$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \varepsilon$
Interaction
$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + ... + \varepsilon$
Quadratic
$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 + ... + \varepsilon$

Figure 4.9: *In DOE, we distinguish between three main types of polynomial models, that is, linear, interaction and quadratic regression models.*

Being the most complex model, a quadratic polynomial requires more experiments than the others. An interaction model requires fewer experiments, and a linear model fewer still. The choice of which model to use is *at this stage* of the research not completely free, because part of the choice was made already in the first step of the problem formulation, that is, when the experimental objective was selected. If optimization was selected, only a quadratic

model will do. If screening was selected, either a linear or an interaction model is pertinent. We recommend an interaction model, if the number of experiments required is practical. If robustness testing was selected, a linear model might be the appropriate choice. The properties of these models (displayed in Figure 4.9) are discussed in Chapter 6.
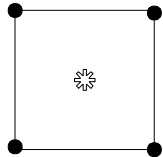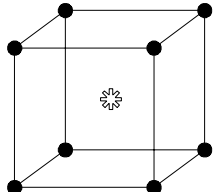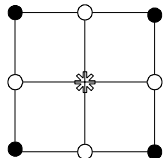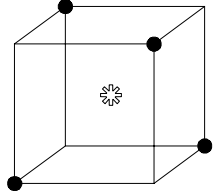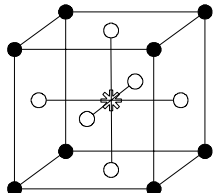
Moreover, one may occasionally end up in a situation where a cubic model is necessary. This is especially the case when modelling the performance of living systems, e.g., growth rate of algae as a function of availability of nutritious agents, water temperature, amount of sunlight, and other factors. However, it should be made clear that cubic models are *rarely* relevant in industrial practice, because they are too complex and demand too many experiments.

Generation of design

Subsequent to the selection of regression model, the next stage in the problem formulation is the generation of an appropriate experimental design. However, one cannot select an arbitrary experimental design and hope that it will work for a given problem. *The chosen model and the design to be generated are intimately linked.* The MODDE software will consider the number of factors, their levels and nature (quantitative, qualitative, …), and the selected experimental objective, and propose a recommended design, which will well suit the given problem. It is possible to override this proposed design, but this choice should only be exercised by experienced DOE-practitioners.

In general, the designs recommended by MODDE will be of the types displayed in Figure 4.10. The upper two rows of the figure show factorial and fractional factorial designs, which are screening designs, and support linear and interaction models. These designs have two investigation levels for each factor, plus an optional number of center-point experiments, here symbolized by the snow-flake. The last row displays composite designs, which are useful for optimization. Composite designs support quadratic models, because every factor will be explored at three or five levels.

2 factors	3 factors	> 3

Figure 4.10: Examples of full factorial, fractional factorial, and composite designs used in DOE.

It is important to observe that these factorial, fractional factorial, and composite designs have a regular geometry, because the experimental region is regular. As soon as the experimental region becomes irregular, for instance, as a result of one corner being inaccessible experimentally, other types of designs displaying irregular geometries have to be employed. The handling of such irregularities in the factor definition and design generation is discussed in Chapter 5.

Creation of experimental worksheet

In the problem formulation, the last stage is the creation of the experimental worksheet. The worksheet is, in principle, very similar to a table containing the selected experimental design. However, in the worksheet it is possible to add vital information linked to the actual execution of the experiments. An example worksheet is shown in Figure 4.11.

	ExpNo	ExpName	RunOrder	InOut	Factor1	Factor2	Factor3	Factor4	Factor5	Response1	Response2
1	1	N1	11	In	10	50	25	25			
2	2	N2	1	In	20	50	25	25			
3	3	N3	5	In	10	100	25	25			
4	4	N4	3	In	20	100	25	25			
5	5	N5	8	In	10	50	75	25			
6	6	N6	2	In	20	50	75	25			
7	7	N7	10	In	10	100	75	25			
8	8	N8	4	In	20	100	75	25			
9	9	N9	6	In	15	75	50	25			
10	10	N10	9	In	15	75	50	25			
11	11	N11	7	In	15	75	50	25			
12	12	N12	12	In	12	80	65	25	25	46	53
13	13	N13	13	In	20	70	65	25	23	49	60
14	14	N14	14	In	15	75	25	25	24	40	89

Figure 4.11: An example worksheet with extra information. Remember to go through the proposed experiments to make sure that all of them look reasonable, can be performed, and have the potential of fulfilling the experimental objective.

One important component in the worksheet is the proposed run order, which tells in which randomized order to conduct the experiments. Another appealing feature is the possibility of incorporating a descriptive name for each trial. Such a name might be linked to a page number of the experimenter's hand written laboratory notebook, or a filename on his computer hard disk, or some other reference system, describing conditions and events occurring during the experimental work.

Furthermore, in the worksheet, one may include additional factors beyond the ones that are actually varied. In the example worksheet, the first three factors are controlled, quantitative and manipulable. We see that the fourth factor is constant. It is not part of the design, but is included in the worksheet to help the experimenter. Also, the uncontrollable fifth factor is included in the worksheet, because it is important to record the values of an uncontrollable factor, and during the data analysis to examine whether this factor had an influence on the responses. Finally, it is possible to append to the worksheet supplementary experiments already carried out. This is done in order to consider such information in the data analysis. In the example worksheet, the last three rows illustrate such existing experimental information.

Quiz II

Please answer the following questions:

What are controllable and uncontrollable factors?

Why is it important to monitor uncontrollable factors?

What are process and mixture factors?

What are the differences between quantitative and qualitative factors?

What is the minimum number of levels which must be defined for a quantitative factor?

How many levels are reasonable for investigating the impact of a qualitative factor?

Why are quantitative responses more useful than qualitative ones?

When is a linear model applicable? Interaction model? Quadratic model?

How many levels of each factor is specified by a composite design?

Which kind of model is supported by a composite design?

Summary

In this chapter, a detailed account of the problem formulation step in DOE was given. The problem formulation is composed of six steps, (i) selection of experimental objective, (ii) specification of factors, (iii) specification of responses, (iv) selection of regression model, (v) generation of design, and (vi) creation of worksheet. To understand which are the relevant experimental objectives, a discussion regarding six stages in DOE was carried out. The six highlighted stages were (i) familiarization, (ii) screening, (iii) finding the optimal region, (iv) optimization, (v) robustness testing, and (vi) mechanistic modelling.

5 Problem formulation (Level 2)

Objective

In this chapter, the aim is to expand the discussion regarding the factor and response definitions in the problem formulation. To this end, we will address qualitative factors at two or more levels, and some complications that arise in the design generation. For instance, with qualitative factors it is no longer possible to define pure center-point experiments, and the balancing of a design is also rendered more difficult. Moreover, we will discuss the situation that arises when it is impossible to make experiments at certain factor combinations, for instance, when parts of the experimental region are inaccessible for experimentation. Simple linear constraints may then be defined for the quantitative factors involved, and used to regulate where the experiments of the design are positioned. Finally, the topic of selecting an appropriate metric for factors and responses will be discussed. Selecting the correct metric involves considering whether a transformation is needed, and which kind of transformation is the optimal one.

Qualitative factors at two levels

A qualitative factor is a categorical variable which is varied in discrete steps, that is, it can only assume certain distinct levels. As an illustration, consider General Example 3 in which the fifth factor, the type of stationary phase in the HPLC system, is a qualitative factor. This factor is said to have two levels, or settings, column type A and type B (see Figure 5.1). Clearly, there is no center level definable in this case, because a hybrid A/B column does not exist. This is graphically illustrated in Figure 5.2, using the factor sub-space defined by the three factors AcN, pH and column type.
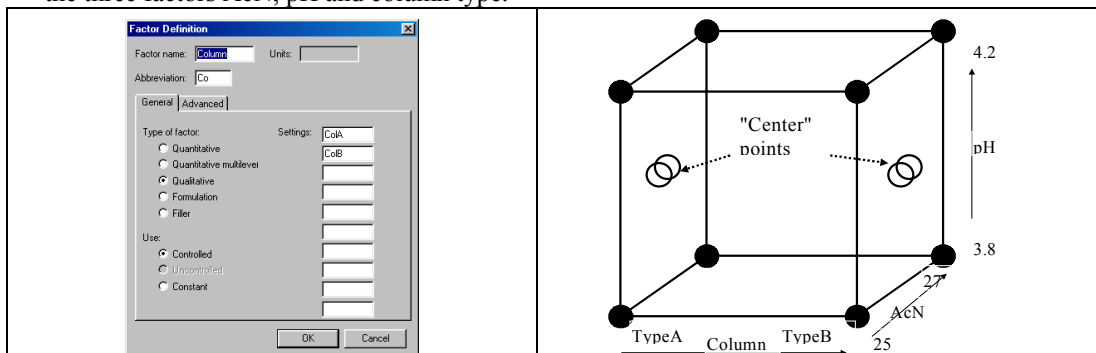
Figure 5.1: *(left) Factor definition of the factor Column of General Example 3.*
Figure 5.2: *(right) A geometrical representation of the design underlying General Example 3.*

All the 12 performed experiments are plotted in Figure 5.2, and we can see that in this three-dimensional subspace they pattern a regular two-level factorial design. Evidently, the experimenter has chosen to make replicated pseudo center-points, located at the centers of the left and right surfaces of the cube. In this way, at least the quantitative factors are varied in three levels, which is desirable when it comes to the data analysis. Another alternative might be to replicate the whole cube once, that is, doing one extra experiment at each corner, but this is more costly in terms of the number of experiments.

Thus, in summary, it is not possible to define technically correct center-points when dealing with two-level qualitative factors, but it is often possible to define centers-of-surfaces which will encode reasonable alternative experiments. In addition, qualitative factors at two levels are easily handled within the standard two-level factorial and fractional factorial framework.

Qualitative factors at many levels

When a qualitative factor is explored at three or more levels, the factor definition is as simple as with two-level qualitative variables. What becomes more challenging, however, is the task of creating a good experimental design with few runs. Full factorial designs with qualitative factors at many levels are unrealistic. Imagine a situation where two qualitative factors and one quantitative factor are varied. Factor A is a qualitative factor with four levels, factor B is also qualitative with three settings, and Factor C a quantitative factor changing between –1 and +1. A full factorial design would in this case correspond to 4*3*2 = 24 experiments, which is depicted by all filled and open circles in Figure 5.3.
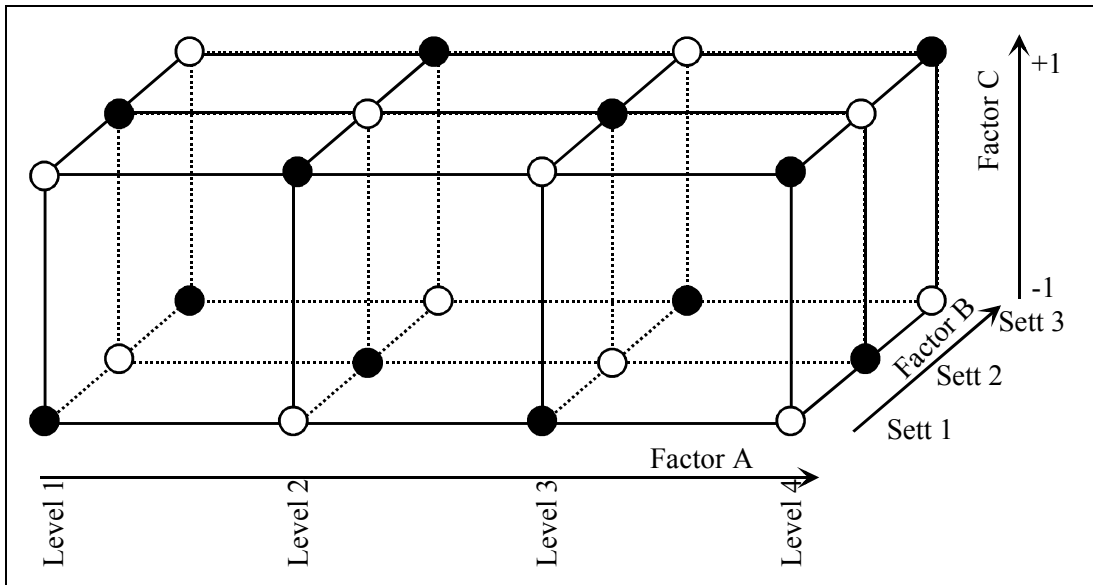
*Figure 5.3: A full 4*3*2 factorial design in 24 experiments.*

With a screening objective and a linear model, 24 runs is unnecessarily high, and an alternative design with fewer experiments is therefore legitimate. Such an alternative design in only 12 experiments is given by the solid circles, or equivalently, by the set of unfilled circles. These subset selections were made using a theoretical algorithm, a D-optimal algorithm, for finding those experiments with the best spread and best balanced distribution. A balanced design has the same number of runs for each level of a qualitative factor. This

feature is not absolutely necessary, but often convenient and makes the design a lot easier to understand and evaluate.

With simple applications and few qualitative factors at few levels, the generation of a suitable design may be done by hand using simple geometrical tools like squares and cubes. However, with more complex applications involving several qualitative factors at many levels, the construction of an appropriate design is by no means an easy task. This latter task demands a D-optimal selection of experiments. D-optimal designs are discussed in Chapter 18.

Qualitative factors in screening, optimization and robustness testing

We now know that qualitative factors at more than two levels require different types of experimental design, other than the regular two-level factorials and fractional factorials. Further, because no center-points are determinable for such factors, replicated experiments cannot be located at the interior part of a design, but must be positioned elsewhere. Also, we would stress that qualitative factors with more than two levels are mainly encountered in screening and robustness testing. This is because in screening the idea is to uncover which setting of a qualitative factor is most favorable, and in robustness testing to reveal how sensitive a response is to changes in such settings. In optimization, however, usually the advantageous level of a qualitative factor will have been found, and there is no need to manipulate such a factor. Thus, most optimization studies will be done by keeping qualitative factors fixed and changing quantitative factors until an optimal point, or a region of decent operability, is detected.

Regular and irregular experimental regions

We will now consider the concept of the *experimental region*, which is an integral part of the problem formulation. The experimental region arises as a result of the factor definition, and corresponds to the factor space in which experiments are going to be performed. It is usually easy to comprehend the geometry of such a region by making plots. The simplest experimental regions are those of regular geometry, such as, squares, cubes, and hypercubes, which are linked to quantitative and qualitative factors without experimental restrictions (Figure 5.4).
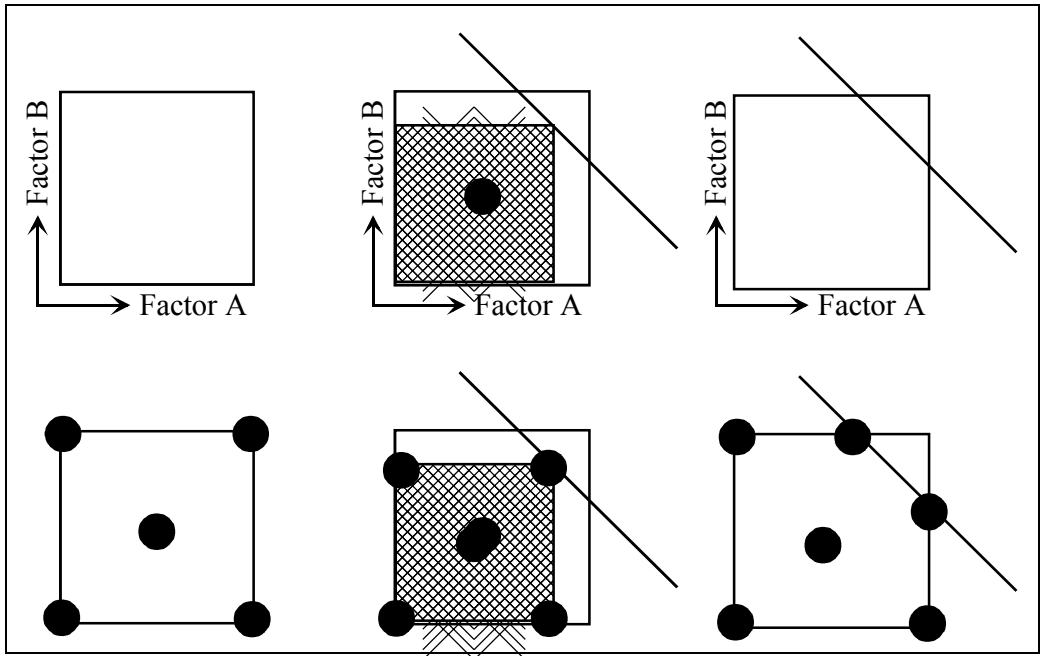
Figure 5.4: *Regular and irregular experimental regions, and some ways to tackle them.*

The left-hand column of Figure 5.4 shows a regular experimental region in two factors, and how it might be mapped with a suitable experimental design. Designs yielding such regular regions are of the factorial, fractional factorial, and composite design types. Sometimes, when the experimenter has a particular requirement concerning the number of experimental runs, other designs like Plackett-Burman and D-optimal designs may be called upon. These are discussed later.

With irregular experimental regions, on the other hand, laying out an appropriate design is not as straightforward. This is exemplified in the middle and right-hand columns of Figure 5.4, where the upper right corner of the experimental region is inaccessible for experimentation. This often happens in industrial practice, e.g., because of unwanted process complications, excessive costs for raw materials, high energy consumption, different process mechanisms, highly toxic by-products, and so on.

In principle, there are two ways of handling an irregular region. The first is to shrink the factor ranges so that the region becomes regular. This is shown by the hatched area in the figure. However, this shrinking is made at the expense of certain parts of the region now being overlooked. The second way is to make use of a D-optimal design, which is well adapted to spreading experiments in an irregular region. This is demonstrated in the right-hand column. Observe that since an irregular region is more complex than a regular one, the former demands more runs than the latter.

We note that in the problem formulation, an irregular experimental region may be defined by specifying linear *constraints,* depicted by the oblique line, among the factors involved. How to specify such factor constraints is discussed below.

Specifying constraints on factors

A common problem in industry is that experimentation may not be allowed in some portions of the experimental region. It may, for example, not be possible in an experiment to have high temperature and simultaneously low pH, as is illustrated in Figure 5.5. In this case, the lower right corner is going to be ignored, and this must be specified in the problem formulation. This is accomplished by specifying a *constraint*, a linear function, of the two quantitative factors Temperature and pH, and this is depicted by the line chopping off the corner. When an appropriate constraint has been defined it ensures that an irregular region will be considered in the subsequent design generation.
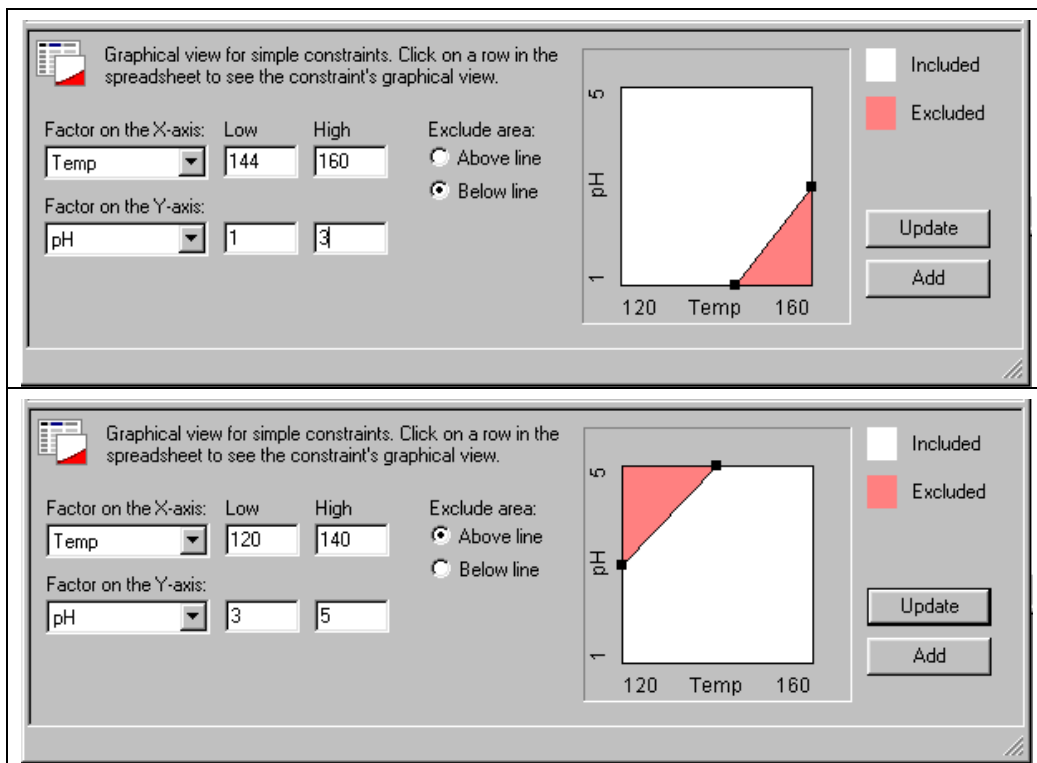
Figure 5.5: (above) Exclusion below line.
Figure 5.6: (below) Exclusion above line.

In more general terms, a constraint may be used to specify which portion of an experimental region that should either be *included* or *excluded*. MODDE includes a functionality for setting simple linear constraints of quantitative factors. Figure 5.5 exemplifies an exclusion below the line, and Figure 5.6 above the line. It is possible to define more than one constraint in an application, and by using the two example constraints one may generate the D-optimal design shown in Figure 5.7.
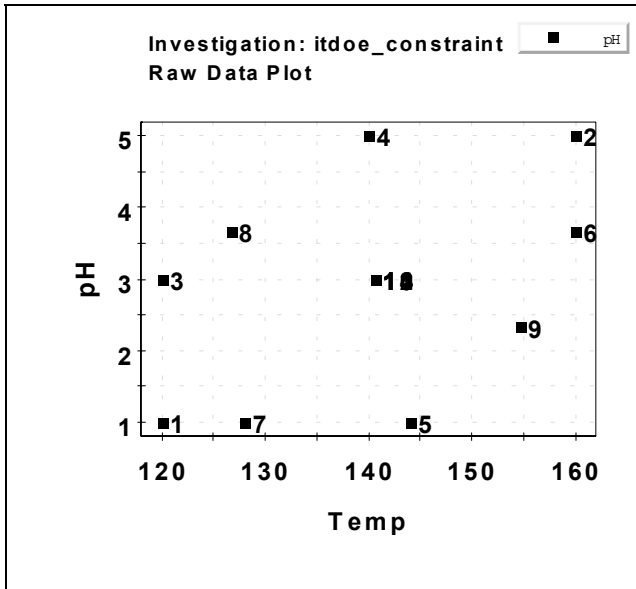
Figure 5.7: *A D-optimal design mapping the irregular region arising from the two linear constraints of Figures 5.5 and 5.6.*

Metric of factors

Another feature regarding the factors that must be treated within the problem formulation is how to set the pertinent *metric* of a factor. The metric of a factor is the numerical system chosen for expressing the measured values, e.g., no transformation or the logarithm transformation. Probably the most well-known transformed factor is pH, the negative logarithm of the H^+ concentration. In fact, pH is one of the five varied factors in General Example 3, the robustness testing study. Another common factor transformation is the square root.

As seen in Figure 5.8, a correctly used factor transformation may simplify a response function, and hence the regression model, by linearizing the non-linearity.
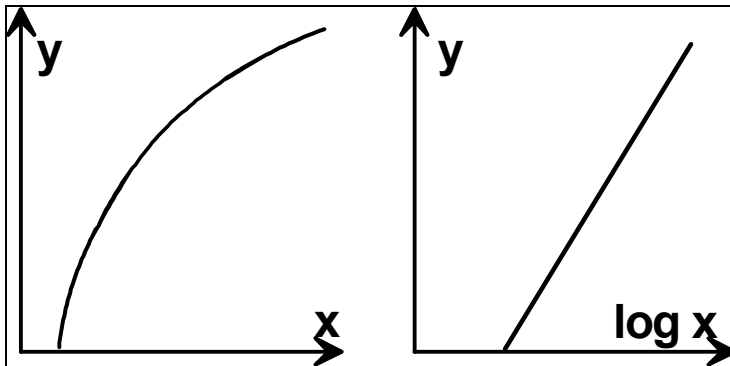
Figure 5.8: *A correctly chosen factor transformation may linearize a response function.*

It is extremely important that a factor transformation is decided upon *before* the worksheet is created and the experiments are executed, because otherwise the design will get distorted. This is evident in the next two figures. The two factors were initially defined as untransformed (Figure 5.9), but in the data analysis of the resulting data, one factor was log-transformed, an operation completely ruining the symmetry of the original design (Figure 5.10).
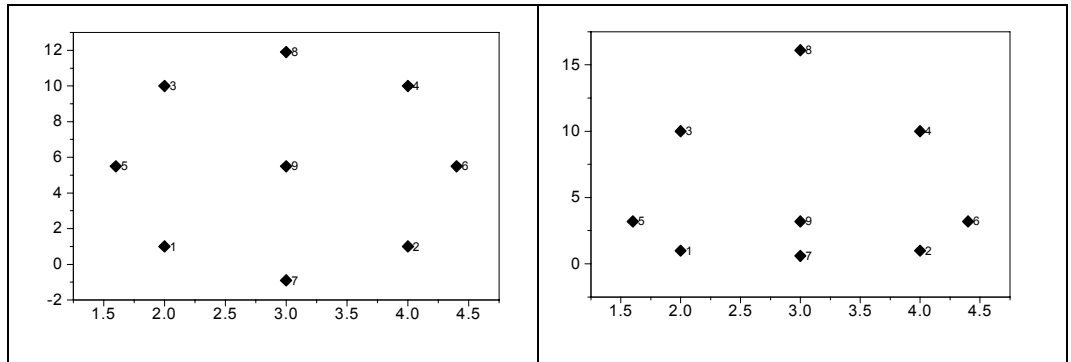
Figure 5.9: *(left) Original design with two untransformed factors.*
Figure 5.10: *(right) Design distortion due to a log-transformation of one of the factors.*

To carry out a factor transformation, one specifies low and high levels as usual, and then the design is created in the transformed unit. For practical purposes, all factor values are then re-expressed in the original unit in the worksheet. Furthermore, it is necessary to ask the question *when is it relevant to transform a factor*? This is not a simple question, but our general advice is that as soon as it is rational to think in powers of 10 for a factor, that factor should probably be log-transformed. Typical examples are concentrations and amounts of raw materials. Other clues may come from the way a factor is expressed, say, a volume in m^3, where the third-root transformation might be useful. Factor transformations are common in analytical and biological systems.

Metric of responses

Contrary to the transformation of factors, the transformation of responses is *not* a critical part of the problem formulation. The transformation of a response does *not* affect the shape and geometry of the selected experimental design, and is always performed in the data analysis stage after the completion of all experiments. A properly selected transformation may (i) linearize and simplify the response/factor relationship, (ii) stabilize the variance, and (iii) remove outliers. The linearization capability is graphically illustrated in Figure 5.11. In Chapters 11 and 12, we will describe a procedure for discovering the need for response transformation. The MODDE software supports seven kinds of response transformations, including the no transformation option.
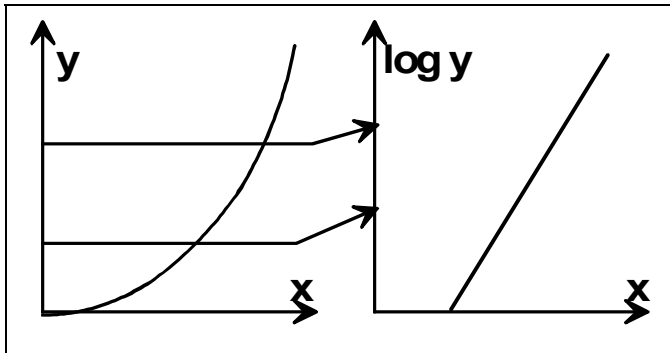
Figure 5.11: A non-linear relationship between y and x, may be linearized by a suitable transformation of y.

Quiz

Please answer the following questions:

Why is it impossible to define replicated center-point experiments when working with qualitative factors?

What are the alternative approaches for creating replicated experiments?

What are the problems to consider when working with qualitative factors at many levels?

What is a balanced design in the context of qualitative factors?

How many levels of qualitative factors are normally investigated in screening? In optimization? In robustness testing?

What is the difference between regular and irregular experimental regions?

When may irregular experimental regions occur?

Which type of design is useful for irregular regions?

What is a constraint?

What is meant by the *metric* of factors and responses?

Which technique is usually used for changing the metric of factors?

Why does the transformation of factors have a decisive impact in the problem formulation?

Summary

This chapter has been devoted to a deeper treatment of factor and response specifications in the problem formulation. Qualitative factors at two or more levels have been addressed, and it has been shown how replicated experiments and balancing of D-optimal designs may be achieved. Further, some attention was given to the experimental region concept, and regular and irregular regions were contrasted. In this context, the concept of linear constraints among quantitative factors was described. Factor constraints are useful for demarcating portions of the experimental region where experiments are undesired. Lastly, parts of this chapter explained the metric of factors and responses, and the need to specify the factor metric *prior* to the design generation was strongly underlined.

6 Problem formulation (Level 3)

Objective

The aim of this chapter is to describe the semi-empirical polynomial models that are commonly chosen in DOE. We will examine some properties of linear, interaction and quadratic models, and pay special attention to their geometric features. We will also review how these models are associated with appropriate experimental designs.

Overview of models

As was discussed in Chapter 4, the selection of a proper regression model is an important part of the problem formulation. It is important to remember that a model is a mathematical representation of the investigated system or process. This model can be viewed as a map which is valid only in the investigated region. It may be risky if a model is used for outside orientation. This was discussed in Chapter 3, where the main emphasis was placed on models as a conceptual phenomenon. In this chapter we will highlight the *geometric* properties of such models.

We recall that with DOE it is possible to establish semi-empirical models, that is, empirical models with some theoretical input, reflecting the underlying true relationship between factors and responses. Linear and interaction models are normally used in screening. Similar models are employed in robustness testing, although linear models are prevalent. Unlike screening and robustness testing, optimization investigations use quadratic models, or RSM models, which are more flexible and adaptable to complex response functions. Other models, such as partly cubic or cubic are sometimes seen in optimization modelling. Regardless of which model is selected for use in DOE, it must be borne in mind that all of them are approximations, simplifications, of a complicated reality. We must avoid being gullible and making over-interpretations of the meaning of a model.

Geometry of linear models

We will start by examining the geometry of linear models and use General Example 3 as an illustration. This example is a robustness testing application, in which five factors are explored in twelve experiments. The experimental plan supports a linear model. In this linear model, each of the five factors occurs as a linear term only. This is shown Figures 6.1 and 6.2.
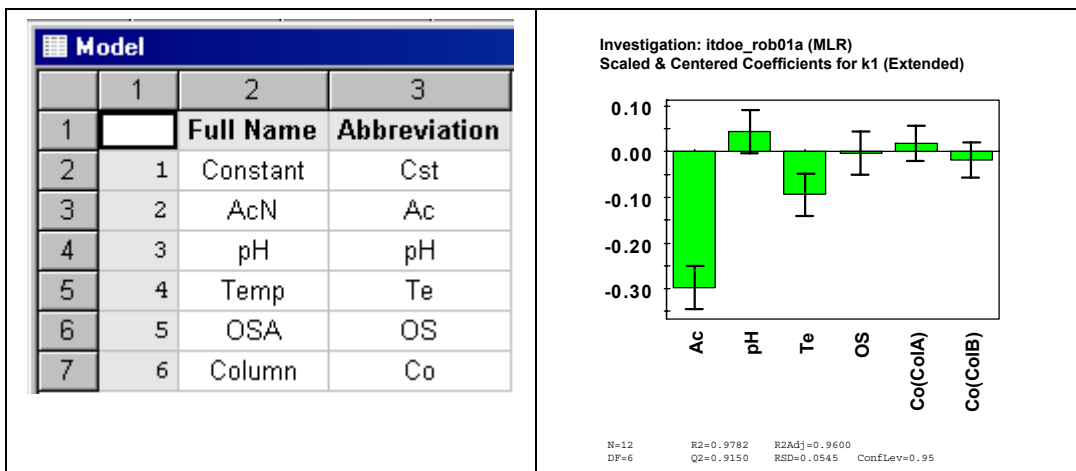
Figure 6.1: (left) The linear model of General Example 3.
Figure 6.2: (right) Regression coefficients of the linear model of k_1.

Let us take a closer look at how two of the five factors, acetonitrile and temperature, influence the first response, the k_1 capacity factor. One illustrative way of doing this is to construct either a response contour plot or a response surface plot of k_1, the latter of which is shown in Figure 6.3. We can see that the linear model forms a *plane* in the space defined by the response and the two factors. With more than two factors, the linear model corresponds to fitting a *hyperplane* in the space defined by the response and the factors x_1, x_2, …., x_k. Another way to inspect a linear model is to make a so called *main effect* plot of each single factor. Such a plot is shown in Figure 6.4 and this graph applies to the acetonitrile factor. This line may be interpreted as the front edge of the surface seen in the surface plot (Figure 6.3). The interpretation of Figure 6.4 suggests that when the amount of acetonitrile in the mobile phase is increased from 25 to 27%, the k_1 capacity factor decreases by almost 0.6 unit, from 2.3 to 1.7.
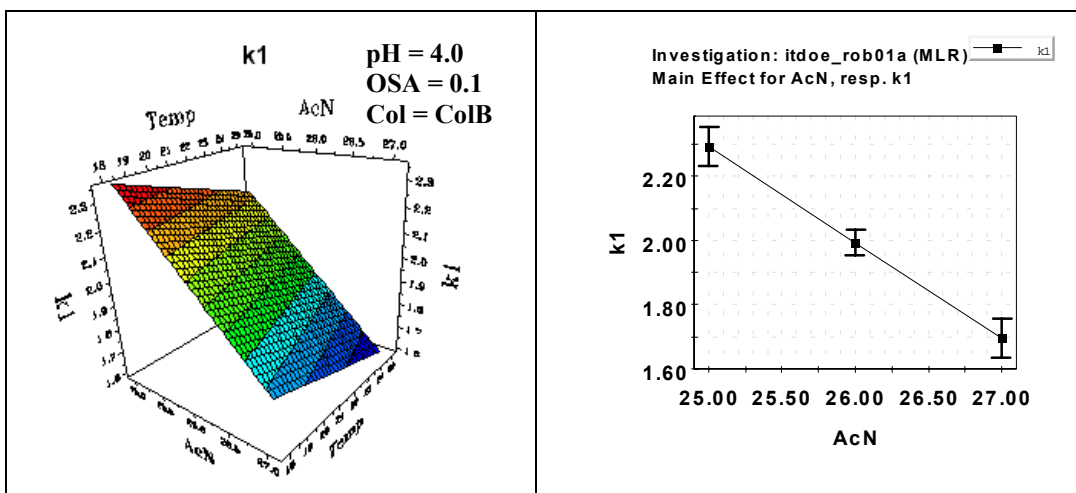
Figure 6.3: (left) Response surface plot of k_1. The varied factors are acetonitrile and temperature.
Figure 6.4: (right) Main effect plot of acetonitrile with regards to k_1.

In addition, there is another feature visible in Figure 6.2, which needs to be highlighted. In this coefficient plot, each quantitative factor is represented by one bar, whereas the impact of the last factor, type of HPLC column, is depicted by two bars. This has to do with the fact that the latter is a qualitative factor in two levels. Since there is no unique center-point for this qualitative factor, one obtains as many bars in the coefficient plot as there are levels of the qualitative factor. The interpretation of the HPLC model indicates that higher k_1 capacity factors are acquired with column type A than type B.

Geometry of interaction models

An interaction model is more complex than a linear model, and may therefore fit more intricate response functions. To understand the geometry of an interaction model, we shall use General Example 1, which is a screening application carried out in two steps. In the first step of this application, the four factors Power, Speed, NozzleGas and RootGas were varied using 11 experiments. The details pertaining to the data analysis are given later in the course material (Chapter 13), and for now we shall focus on the second response, the Width of the weld. It was found that one of the main effects, NozzleGas, had no influence on Width. Further, only one two-factor interaction, the Power*Speed term, was meaningful. Hence, the model obtained is not a full interaction model, but a partial interaction model. Figure 6.5 shows the model composition and Figure 6.6 the model shape.
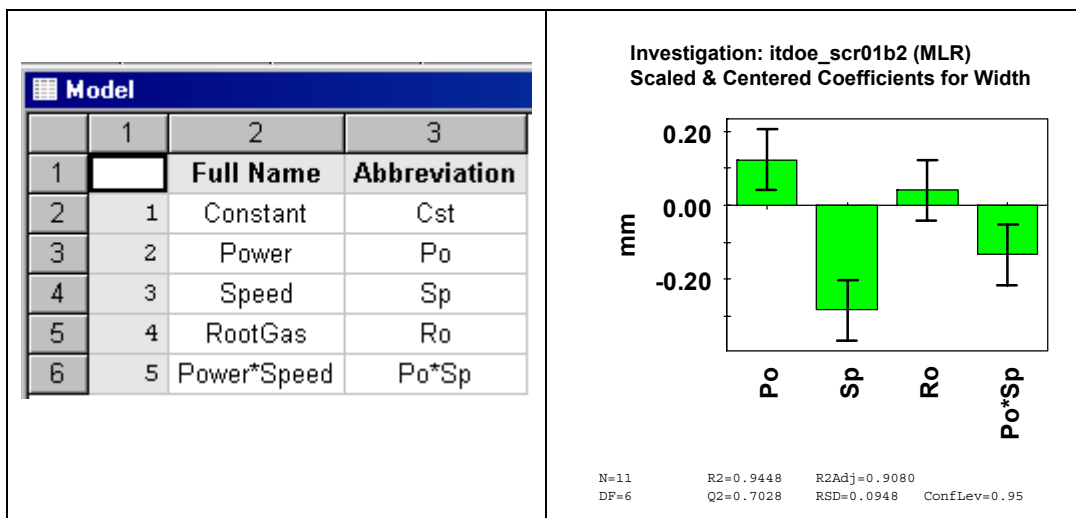
Figure 6.5: *(left) The interaction model of General Example 1.*
Figure 6.6: *(right) Regression coefficients of the interaction model of Width.*

Unlike in a linear model, the surface in the response surface plot of an interaction model (Figure 6.7) is no longer an undistorted plane but a *twisted* plane. Because of the presence of the two-factor interaction, the surface is twisted. This two-factor interaction may be further explored by means of the interaction plot (Figure 6.8), which may be thought of as displaying the front and rear edges of the surface in Figure 6.7. We can see that when changing Speed from low to high level, we get a larger impact on the response when the second factor Power is at its high level rather than its low level. This is the verbal definition of a two-factor interaction: *the effect of one factor depends on the level of the other factor.*
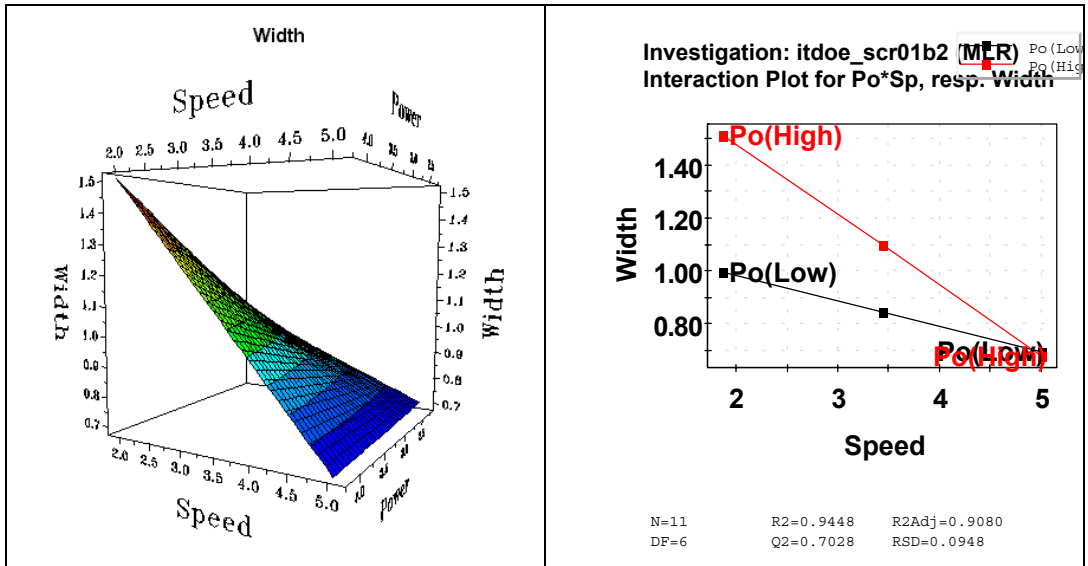
Figure 6.7: *(left) The twisted response surface of Width. The varied factors are Power and Speed.*
Figure 6.8: *(right) Interaction effect plot of Power and Speed with regards to Width.*

Geometry of quadratic models

In order to comprehend the geometry of a quadratic model we shall use General Example 2. This is an optimization investigation of the combustion in a truck engine. Three factors were varied according to a composite design incorporating 17 runs. This design supports a full quadratic polynomial model. Figure 6.9 shows the best model obtained when simultaneously considering the three responses Fuel, NOx, and Soot. Figure 6.10 displays the model appearance with regards to the Fuel response. Apparently the factors Air and NeedleLift dominate the model with regard to Fuel, and a rather large squared term for Air is detected.
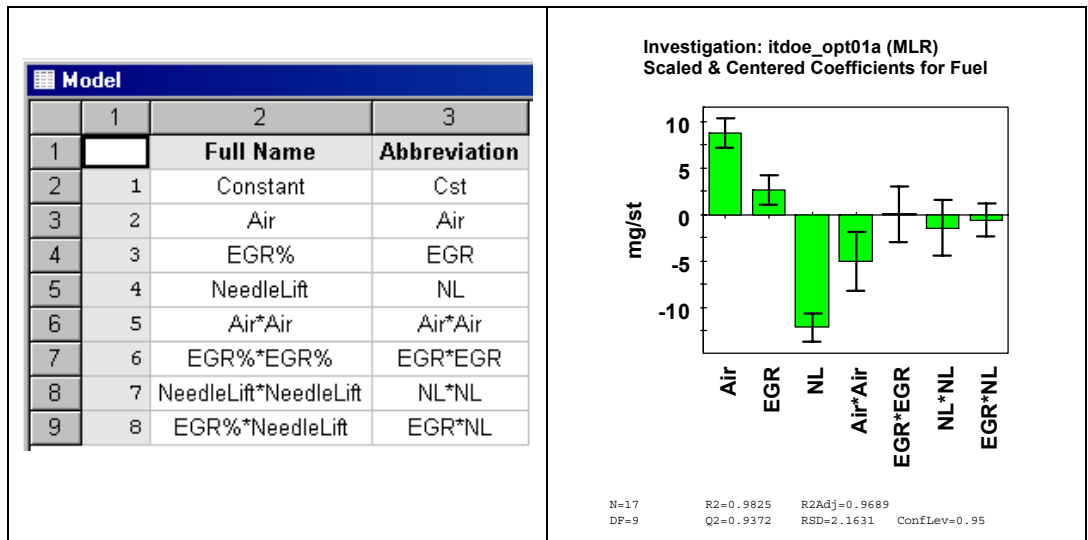
Figure 6.9: *(left) The quadratic model of General Example 2.*
Figure 6.10: *(right) Regression coefficients of the quadratic model of Fuel.*

Further, Figure 6.11 shows the response surface plot of Fuel obtained when varying Air and NeedleLift. This surface is now *curved*, and the possibility of modelling curvature is due to the presence of the quadratic terms. In order to diminish the volume of consumed fuel, the truck engine should be run with a high value of NeedleLift and a low value of Air, according to the model interpretation. Another way of examining the model curvature is to make main effect plots of the separate factors. One such plot, pertaining to the combination Air and Fuel, is provided in Figure 6.12. This plot also indicates that in order to decrease Fuel, the amount of Air used in combustion ought to be lowered.
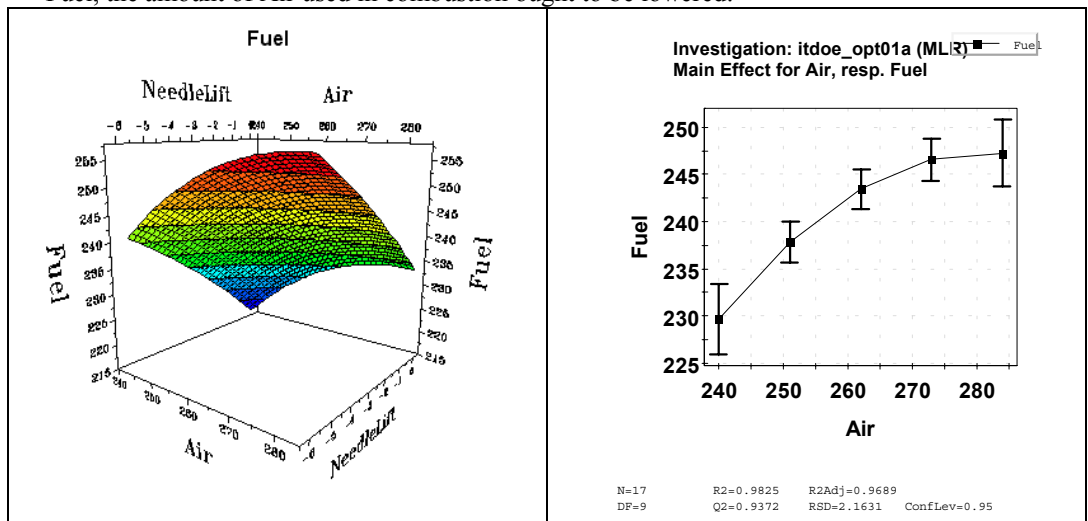
Figure 6.11: *(left) The curved response surface of Fuel. The varied factors are NeedleLift and Air.*
Figure 6.12: *(right) Main effect plot of Air with regard to Fuel.*

Generation of adequate designs

In the problem formulation it is often a question of making an "educated guess" when selecting an appropriate model. As a rule-of-thumb, one can select a linear model for robustness testing, a linear or an interaction model for screening, and a quadratic model for optimization. This model selection then strongly influences the design selection. It is instructive to overview how models and designs are linked, and this is graphically summarized in Figure 6.13.
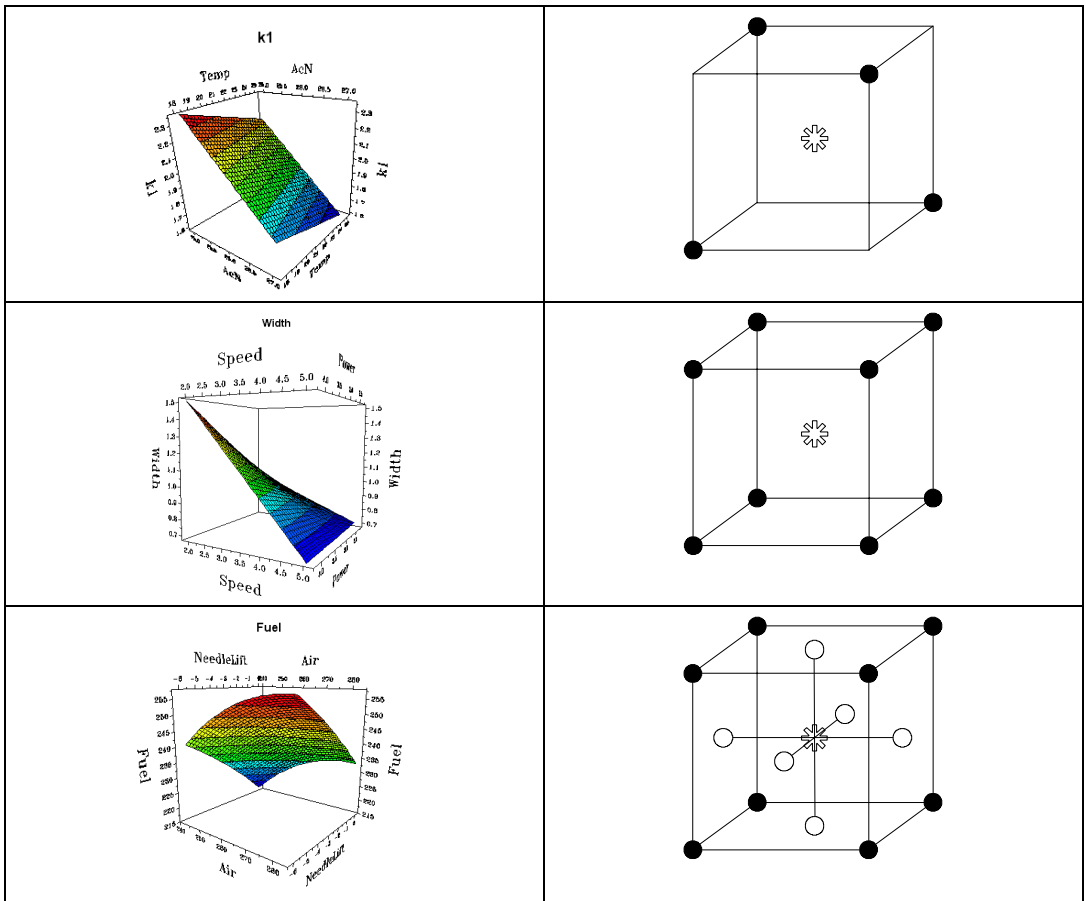
Figure 6.13: An overview of how models and experimental designs are connected to each other.

The first row of Figure 6.13 shows the *plane* surface of the linear robustness testing model, and a simplified version of the corresponding underlying fractional factorial design. For display purposes, the design is drawn in only three factors, although the application itself comprises five factors. The fractional factorial design encodes only a limited number of experiments, because not all corners are investigated. In the second row of the figure, the *twisted* plane of the interaction model of the screening application is shown, together with a complete factorial design in three factors. We can see that an interaction model requires more experiments than a linear model. As far as interaction models are concerned, they may sometimes also be supported by designs drawn from the fractional factorials family. This is

actually the case in the current screening example. Finally, the third row of Figure 6.13 displays the relationship between the *curved* surface of the optimization application and the underlying composite design. The quadratic optimization model demands most experiments.

In reality, it is the data analysis that will reveal whether the initial "educated guess" of model was appropriate. Should it turn out that the experimenter did underestimate the model complexity, it is possible to modify the model and augment the design. For instance, the plotted fractional factorial design may be expanded to the full factorial design, which, in turn, may be augmented to the composite design. With few factors the expansion of a design is easy, with many factors more taxing. In this context, one may talk about *updating* and *upgrading* of models. By updating we mean the addition of a *limited number* of well-identified model terms, for instance, a two-factor interaction term to a linear model. By upgrading we mean the addition of a *set* of model terms, converting, for instance, a linear model to a quadratic model. This will be discussed later (see Chapter 18).

Quiz

Please answer the following questions:

What is the geometric property of a linear model?

What is the geometric property of an interaction model?

What is the geometric property of a quadratic model?

Which type of design is typically linked to a linear model? Interaction model? Quadratic model?

Summary

In this chapter, we have scrutinized the geometric properties of linear, interaction and quadratic polynomial models, which are the kind of models typically used in DOE. A linear model needs a fractional factorial design, an interaction model either a full or a fractional factorial design, and a quadratic model a composite design. When making use of response surface plotting, a linear model corresponds to an undistorted flat or inclined plane, an interaction model to a twisted plane, and a quadratic model to a curved surface.

7 Factorial designs (Level 1)

Objective

Full factorial designs form the basis for all classical experimental designs used in screening, optimization, and robustness testing. Hence, a good understanding of these designs is important, because this will make it easier to understand other related, and more commonly used, experimental designs. In this chapter we will introduce two-level factorial designs. The aim is to consider their use and construction.

Introduction to full factorial designs

In this chapter, we shall consider two-level full factorial designs. Such factorial designs support interaction models and are used in screening. They are important for a number of reasons:

- they require relatively few runs per investigated factor

- they can be upgraded to form composite designs, which are used in optimization

- they form the basis for two-level *fractional* factorial designs, which are of great practical value at an early stage of a project

- they are easily interpreted by using common sense and elementary arithmetic

Factorial designs are regularly used with 2-4 factors, but with 5 or more factors the number of experiments required tends to be too demanding. Hence, when many factors are screened, fractional factorial designs constitute a more appealing alternative.

Notation

To perform a general two-level full factorial design, the investigator has to assign a low level and a high level to each factor. These settings are then used to construct an orthogonal array of experiments. There are some common notations in use to represent such factor settings. Usually, the low level of a factor is denoted by –1 or just –, and the high level by +1 or simply +. As a consequence, the center level, usually chosen for replication, will be denoted by 0. As seen in Figure 7.1, these alternatives are called standard and extended notation. Both these notations are said to operate in a coded, -1 to +1, unit.

Item	Low	High	Center
Standard notation	−	+	0
Extended notation	−1	+1	0
Example; temperature	100ºC	200ºC	150ºC
Example; pH	7	9	8
Example; Catalyst (A, B)	A	B	N/A

Figure 7.1: *Standard and extended notations for factor settings.*

For simple systems, such as the two factor situation sketched in Figure 7.2, it may be convenient to display the coded unit together will the original factor unit. In this example, we easily see that when using catalyst A and raising the temperature from 100ºC to 200ºC, the yield is enhanced.
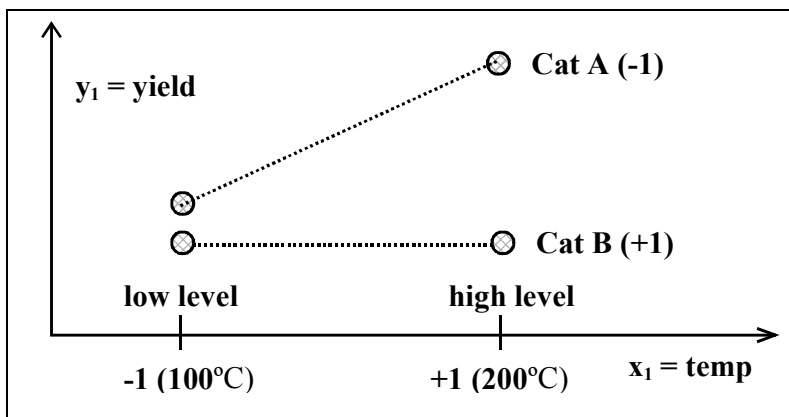
Figure 7.2: *Graphical display of a two-factor example.*
1. The low level of a factor will be denoted by -1 or just -
2. The high level of a factor will be denoted by +1 or just +
3. Consequently, the center is denoted by 0.

The 2^2 full factorial design – construction & geometry

The 2^2 full factorial design is the simplest of its kind, and the 2^2 nomenclature is understood as a *two-level* design in *two* factors. We shall consider a simple example, known as "ByHand", to illustrate the principles of a two-level factorial design. This example stems from the field of organic chemistry and deals with the reduction of an enamine. Enamines are reduced by formic acid to saturated amines. The experimenter decided to vary two factors (Figure 7.3). One factor, x_1, was the molar ratio of the two reacting compounds, formic acid and enamine. As seen in Figure 7.3, this ratio was varied between 1 and 1.5. The second factor, x_2, was the reaction temperature, which was varied between 25ºC and 100ºC. In order to monitor the success of the reaction, three responses were measured. We will only focus on one response, here called y_3, the formation of the desired product, which should be maximized (Figure 7.3).

Factors		Levels		
		- (1)	0	+ (1)
x_1	Amount formic acid/enamine (mole/mole)	1.0	1.25	1.5
x_2	Reaction temperature (ºC)	25	62.5	100
Response				
y_3	The desired product %			

Figure 7.3: *The two factors and the response considered in the ByHand example.*

With two factors and two levels of each, there are four possible factor-combinations, that is, low-low, high-low, low-high, and high-high, which correspond to the four first rows of the design table shown in Figure 7.4. In addition, three replicated experiments carried out at the center of the experimental region have been added. Such experiments are typically termed *center-points*, since they are located midway between the low and high levels.

	Factors		Factors		Response
	Original unit		Coded unit		%
Exp. no	x_1	x_2	x_1	x_2	y_3
1	1	25	-	-	80.4
2	1.5	25	+	-	72.4
3	1	100	-	+	94.4
4	1.5	100	+	+	90.6
5	1.25	62.5	0	0	84.5
6	1.25	62.5	0	0	85.2
7	1.25	62.5	0	0	83.8

Figure 7.4: *The 2^2 factorial design of the ByHand example.*

Geometrically, the experimental design created may be interpreted as a square, and hence the experimental region is said to be of regular geometry (Figure 7.5). The important point is that each row in the experimental design (Figure 7.4) corresponds to one experiment, and may be interpreted as a point in the two-dimensional factor space (Figure 7.5).
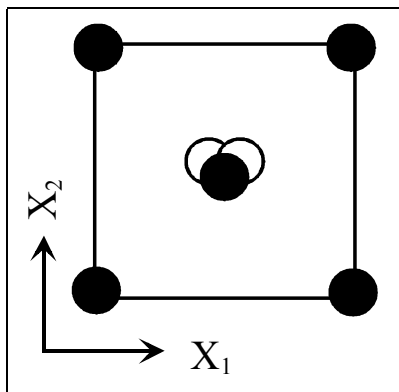
Figure 7.5: *A geometrical representation of the factorial design used in the ByHand example.*

The 2³ full factorial design – construction & geometry

The two-level full factorial design in three factors, denoted 2^3, is constructed analogously to the full factorial design in two factors. As an illustration, we shall use the CakeMix application. In the construction of the 2^3 design, the experimenter first has to assign to each investigated factor a low and a high level. Then the actual *design matrix* is created as follows (Figure 7.6). The first and leftmost column of the design matrix is created by writing minus and plus signs alternatingly in 8 rows, simply because the factorial part of the design will have $2^3 = 8$ rows. Subsequently, the second column is created by successive pairs of minus and plus signs continuing until the first eight rows of the table have entries. Finally, the third column of the design matrix is generated by doubling the number of signs in a sequence, and then alternately laying out such sequences. This implies that the third column starts with four minus signs followed by four plus signs. The design that we have just created is said to be written in *standard order*.

Design Matrix				Experimental matrix			
Exp No	Flour	Shortening	Egg	Flour	Shortening	Egg	Taste
1	-	-	-	200	50	50	3.52
2	+	-	-	400	50	50	3.66
3	-	+	-	200	100	50	4.74
4	+	+	-	400	100	50	5.2
5	-	-	+	200	50	100	5.38
6	+	-	+	400	50	100	5.9
7	-	+	+	200	100	100	4.36
8	+	+	+	400	100	100	4.86
9	0	0	0	300	75	75	4.68
10	0	0	0	300	75	75	4.73
11	0	0	0	300	75	75	4.61

Figure 7.6: *The 2^3 factorial design of the CakeMix example.*

In addition to the eight experiments of the factorial part, it is recommended to incorporate replicated experiments, usually carried out as center-points. In the CakeMix case, three center-points were added, denoted with zeroes in the design matrix (rows 9-11 in Figure 7.6). In order to facilitate the experimental work, the *design matrix* may also be converted to the *experimental matrix* by inserting the original factors settings. This is also shown in Figure 7.6. Every row in the design table, that is, each experiment, represents a point in the three-dimensional experimental space. Obviously, the experimental region in the CakeMix case is a cube of regular geometry (Figure 7.7).
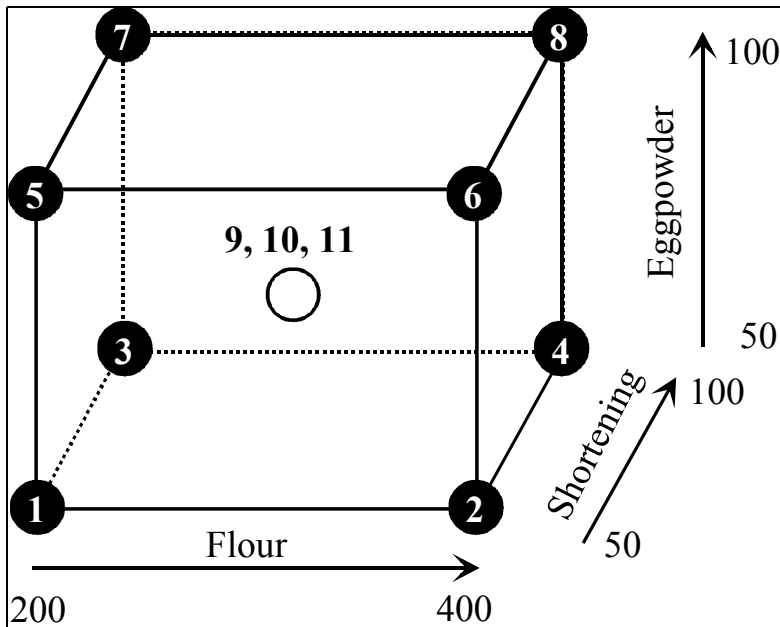
Figure 7.7: The three-dimensional experimental region of the CakeMix example.

The 2^4 and 2^5 full factorial designs

We will now consider the construction of the 2^4 and 2^5 full factorial designs. Fortunately, the procedure described in the foregoing paragraph is useful for this purpose. First we consider the 2^4 design. The trick is to compute the number of necessary rows, which in this case is 2^4 = 16, and then lay out the first and leftmost column by means of a series of alternating minus and plus signs (Figure 7.8). To complete the second and third columns we adhere to the already described procedure and fill them with 16 entries. Finally, the fourth column is created by first writing eight minus signs and then eight plus signs. This design matrix is indicated by the grayed area in Figure 7.8.

	X 1	X 2	X 3	X 4	X 5
1	-	-	-	-	-
2	+	-	-	-	-
3	-	+	-	-	-
4	+	+	-	-	-
5	-	-	+	-	-
6	+	-	+	-	-
7	-	+	+	-	-
8	+	+	+	-	-
9	-	-	-	+	-
1 0	+	-	-	+	-
1 1	-	+	-	+	-
1 2	+	+	-	+	-
1 3	-	-	+	+	-
1 4	+	-	+	+	-
1 5	-	+	+	+	-
1 6	+	+	+	+	-
1 7	-	-	-	-	+
1 8	+	-	-	-	+
1 9	-	+	-	-	+
2 0	+	+	-	-	+
2 1	-	-	+	-	+
2 2	+	-	+	-	+
2 3	-	+	+	-	+
2 4	+	+	+	-	+
2 5	-	-	-	+	+
2 6	+	-	-	+	+
2 7	-	+	-	+	+
2 8	+	+	-	+	+
2 9	-	-	+	+	+
3 0	+	-	+	+	+
3 1	-	+	+	+	+
3 2	+	+	+	+	+

Figure 7.8: *The 2^4 (grayed area) and 2^5 factorial designs.*

Geometrically, the 2^4 design corresponds to a regular hypercube with four dimensions. In fact, the laser welding application may be used as an illustration of the 2^4 design. Although this application was carried out in two steps, the combined factorial parts form the 2^4 design. This is shown in Figure 7.9, but note that for clarity the performed center-points have been omitted.

Figure 7.9: *The combined factorial parts of the laser welding experiments form a 2^4 factorial design.*

Further, the 2^5 design is constructed similarly to the 2^4 design, except that the design matrix now has 32 rows and 5 columns. This design matrix is shown in Figure 7.8, and geometrically corresponds to a regular five-dimensional hypercube. The 2^5 full factorial design is not used to any great extent in industrial practice, because of the large number of experimental trials. Instead, there exists an efficient fractional factorial design in 16 runs, which is almost as good as the full factorial counterpart. Also observe that to the 2^4 and 2^5 designs 3 to 5 center-points are normally added.

Pros and cons of two-level full factorial designs

With two-level full factorial designs it is possible to estimate interaction models, which well serve to fulfil the objectives underlying screening. This type of design consists of a *set* of experimental runs in which each factor is investigated at both levels of all the other factors. In other words, the two-level full factorial design corresponds to a *balanced* and *orthogonal* arrangement of experiments. This arrangement enables the effect of one factor to be assessed independently of all the other factors. With k investigated factors the two-level full factorial design has $N = 2^k$ runs. In Figure 7.10, we have tabulated the number of runs for k = 2 to k = 10. It must be observed that this compilation does *not* take any replicated experiments into account.

No of investigated factors (k)	No of runs Full factorial	No of runs Fractional factorial
2	4	---
3	8	4
4	16	8
5	32	16
6	64	16
7	128	16
8	256	16
9	512	32
10	1024	32

Figure 7.10: *An overview of the number of experiments encoded by two-level full and fractional factorial designs in k = 2 to k = 10 factors.*

As seen in Figure 7.10, with more than 5 factors the number of experiments increases dramatically, so that only the full factorials with 2 to 4 factors are realistic choices. The rightmost column lists the more manageable number of experiments required by designs of the two-level fractional factorial family. For example, it is realistic to screen 10 factors in 32 experiments, plus some optional replicates. In summary, two-level full factorial designs are experimentally practical and economically defendable only when considering few factors. With more than four factors, a switch to fractional factorial designs is more favorable.

Quiz

Please answer the following questions:

What is a two-level factorial design?

How many runs are included in the factorial part of the 2^2 design? 2^3? 2^4? 2^5?

What is the geometry of the 2^3 design?

How many center-point experiments are usually incorporated in a factorial design?

How many factors are practical for two-level full factorial designs?

When is it recommended to switch to fractional factorial designs?

Summary

In this chapter, we have provided an overview of two-level full factorial designs. Such designs support interaction models and are used in screening. Full factorial designs are balanced and orthogonal, which means that the influence of one factor on the result may be determined independently of all other studied factors. Full factorial designs are useful for up to four factors. When more than four factors are screened, fractional factorial designs are more tractable. We have also considered the construction and geometry of two-level full factorial designs.

8 Factorial designs (Level 2)

Objective

The aim of this chapter is to define the main effect of a factor and the interaction effect between two factors, and to illustrate how these may be displayed graphically. An additional aim is to elucidate some elementary ways of analyzing data from full factorials by means of simple arithmetic. This is done in order to increase the understanding of basic concepts. We will also briefly discuss the relationship between effects and regression coefficients.

Main effect of a factor

We will now describe the main effect of a factor. Consider the graph in Figure 8.1. This is a graphical representation of how a response, y_1, may change due to changing values of the factor, x_1. The response might be the taste of a cake as in the CakeMix case, and the factor the amount of flour consumed. The main effect of a factor is defined as the change in the response due to varying one factor from its low level to its high level, and keeping the other factors at their average. Thus, in our case, the main effect of flour is the change in taste when increasing the amount of flour in the cake mix recipe from 200g to 400g, and keeping shortening and egg powder at their center level of 75g.
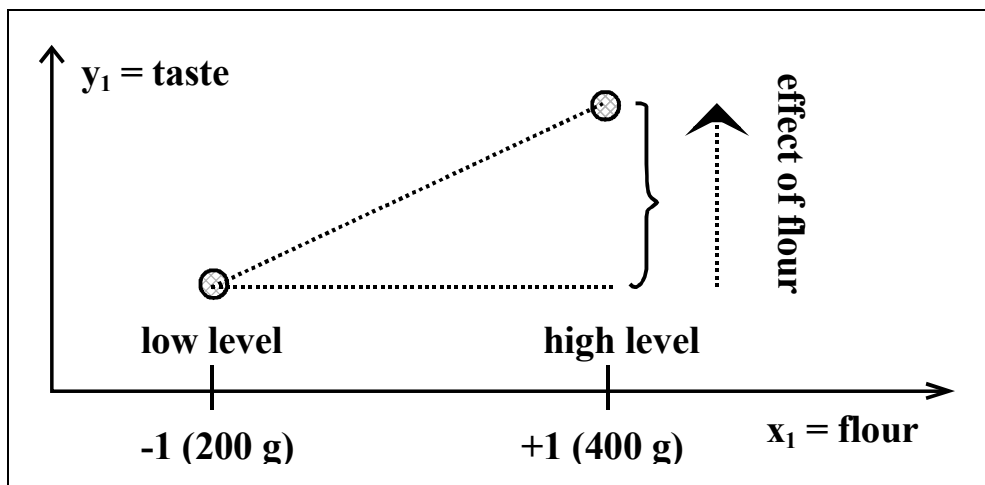
Figure 8.1: Illustration of the main effect of a factor, e.g., flour in cake mix recipe.

Moreover, we can see in Figure 8.2 that the main effect of flour amounts to approximately 0.4 unit, from taste 4.5 to 4.9. Since the experimental plan in the CakeMix case is a two-

level full factorial design, all factors are varied simultaneously and in a systematically balanced manner. This allows the estimation of a factor's main effect independently of the other factors.
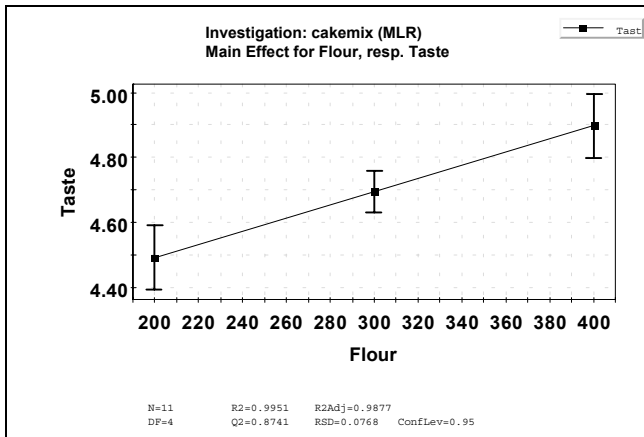
Figure 8.2: *Main effect plot of flour with regard to taste.*

Computation of main effects in the 2^2 case

With a 2^2 factorial design two main effects may be calculated. To illustrate graphically and arithmetically how this is carried out, we shall revisit the ByHand example introduced in Chapter 7. Recall that this example has two factors, the molar ratio of formic acid/enamine and the reaction temperature, and that focus was given to the response reflecting the yield of the desired product. We may ask ourselves, *what is the main effect of the molar ratio on the yield of the desired product*? One manner in which to compute and geometrically understand this main effect is given in Figure 8.3. There are two estimates of the impact of changing the molar ratio from 1.0 to 1.5. One estimate, called Δ1,2, indicates that the yield will decrease by 8%. Another estimate, denoted Δ3,4, suggests that the yield will drop only 3.8%. To compute the *main effect* of the molar ratio we simply average these two estimates, that is, the main effect of the molar ratio is –5.9%, which is plotted in Figure 8.4.
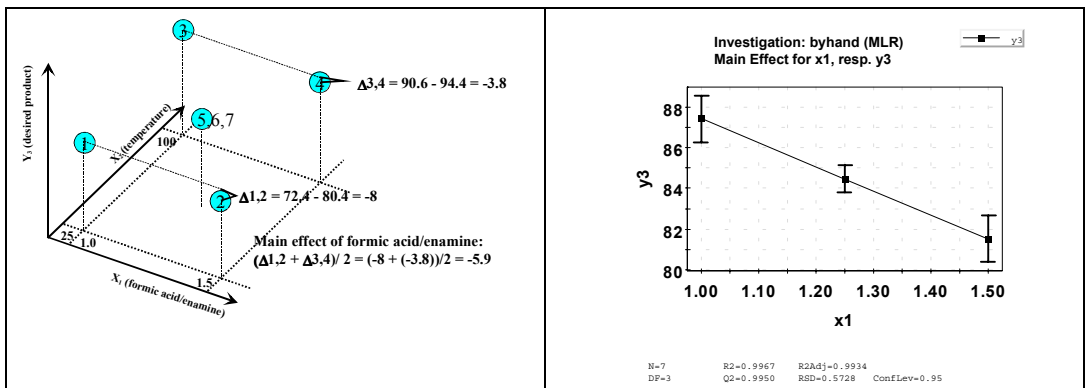
Figure 8.3: *(left) Graphical illustration of the main effect of molar ratio.*
Figure 8.4: *(right) Main effect plot of molar ratio.*

If we want to compute the main effect of the other factor, the reaction temperature, we proceed similarly but along the temperature axis. This is portrayed in Figure 8.5. We can see that the two estimates, called Δ1,3 and Δ2,4, of the temperature effect, are 14.0 and 18.2, respectively. The average of these two values, 16.1, corresponds to the main effect of the temperature, that is, the change in yield when raising the temperature from 25ºC to 100ºC. This main effect is plotted in Figure 8.6.
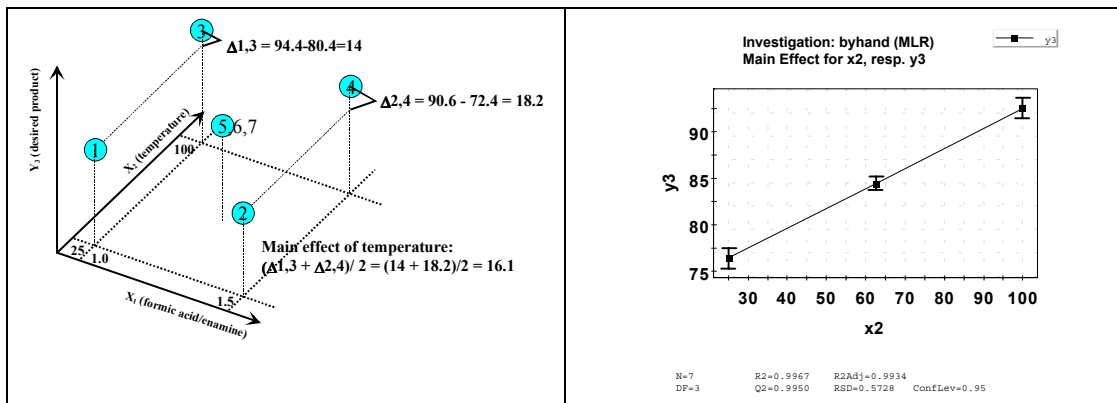
Figure 8.5: (left) Graphical illustration of the main effect of temperature.
Figure 8.6: (right) Main effect plot of temperature.

A second method of understanding main effects

A second way of understanding and computing the main effects of the molar ratio and the reaction temperature is to make pair-wise comparisons of the edges of the design square. Figure 8.7 shows how this may be done to derive the main effect of the molar ratio. By forming the average of the right side, 81.5, and subtracting from this value the average of the left side, 87.4, one obtains −5.9, which is the estimate of the main effect of the molar ratio. Compare with the main effect plotted in Figure 8.4. Analogously, the main effect of the reaction temperature may be formed by comparing the rear edge with the front edge. As seen in Figure 8.8, the rear edge average is 92.5 and the front edge average 76.4, which gives the main effect of 16.1. Compare with the temperature main effect plotted in Figure 8.6. Notice (1) that all four corner experiments are used to supply information on each of the main effects, and (2) that each effect is determined with the precision of a two-fold replicated difference.
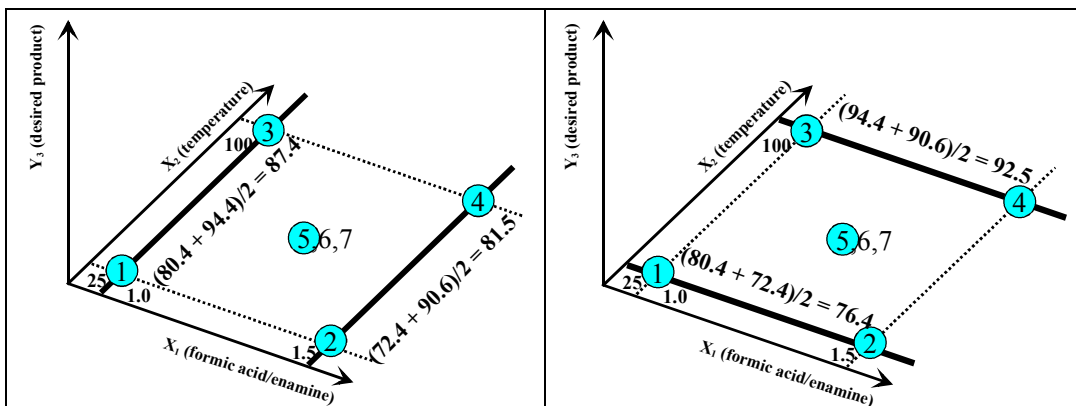
Figure 8.7: (left) Alternative method for computing the main effect of molar ratio.
Figure 8.8: (right) Alternative method for computing the main effect of reaction temperature.

A quicker by-hand method for computing effects

We will now examine a quicker method for computing the effects of factors, which is based on using the columns of the computational matrix. Figure 8.9 gives the computational matrix of the ByHand example.

	Experimental matrix		Computational matrix				Response
Exp. no	x1	x2	mean	x1	x2	x1*x2	y3
1	1	25	+	-	-	+	80.4
2	1.5	25	+	+	-	-	72.4
3	1	100	+	-	+	-	94.4
4	1.5	100	+	+	+	+	90.6
5	1.25	62.5	+	0	0	0	84.5
6	1.25	62.5	+	0	0	0	85.2
7	1.25	62.5	+	0	0	0	83.8

Figure 8.9: Experimental and computational matrices of ByHand example. Calculations below refer to the computational matrix.
1st column gives the mean: $(+80.4+72.4+94.4+90.6+84.5+85.2+83.8)/7 = 84.5$;
2nd column gives the molar ratio, x_1, main effect: $(-80.4+72.4-94.4+90.6)/2 = -5.9$;
3rd column gives the reaction temperature, x_2, main effect: $(-80.4-72.4+94.4+90.6)/2 = 16.1$;
4th column gives the x_1*x_2 two-factor interaction: $(+80.4-72.4-94.4+90.6)/2 = 2.1$

The first column of the computational matrix does not provide any information related to factor effects, but is used to compute the average response according to $(+80.4+72.4+94.4+90.6+84.5+85.2+83.8)/7 = 84.5$. The molar ratio, x_1, main effect is calculated from the second column according to $(-80.4+72.4-94.4+90.6)/2 = -5.9$. Observe that the replicated center-point rows of the computational matrix do not contribute to this computation. Analogously, the reaction temperature, x_2, main effect is calculated according to $(-80.4-72.4+94.4+90.6)/2 = 16.1$. Finally, the fourth column, derived by multiplying the x_1 and x_2 columns, is used to encode the two-factor interaction molar ratio*reaction

temperature, x_1*x_2, according to $(+80.4-72.4-94.4+90.6)/2 = 2.1$. Interestingly, by using the same graphical interpretation as in the preceding paragraph, this two-factor interaction may be interpreted as the difference between the response averages of the black and gray diagonals in Figure 8.10, that is, $85.5-83.4 = 2.1$.
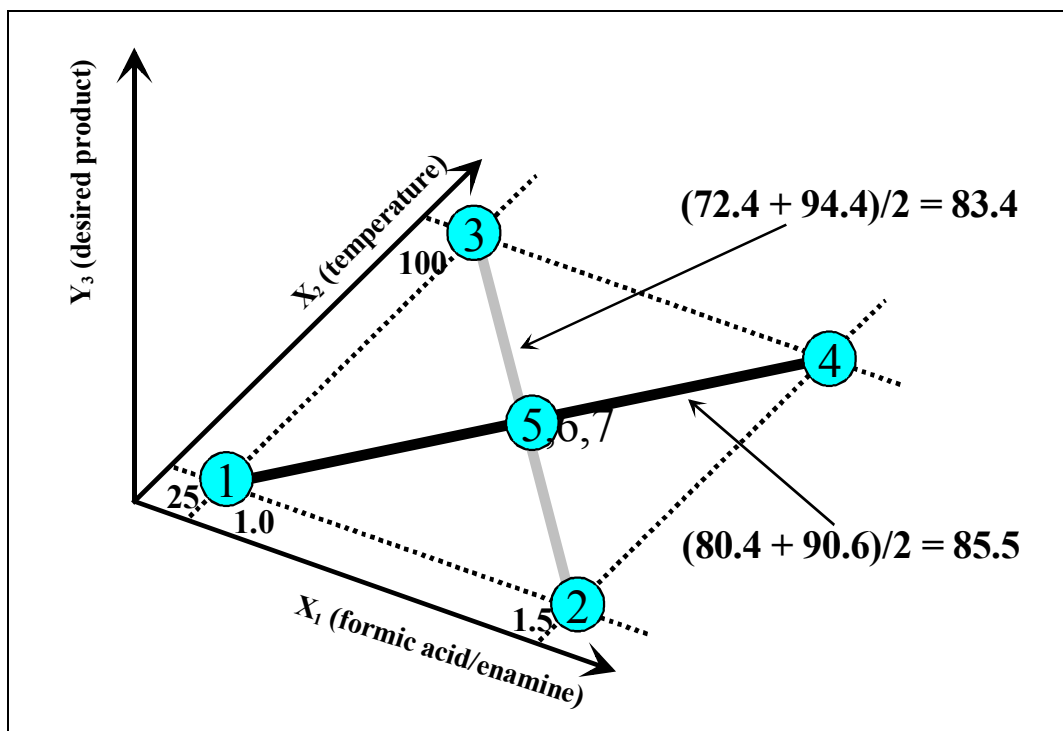
Figure 8.10: The molar ratio*reaction temperature two factor interaction is interpreted as the difference between the black and gray diagonal averages.

Plotting of main and interaction effects

Main effects and two-factor interaction effects have different impacts on the appearance of a semi-empirical model. Figure 8.11 shows the response surface obtained when fitting an interaction model to the yield of the desired product. The two main effects make the surface slope and the two-factor interaction causes it to twist. This is one way of interpreting these effects.

In addition, it is possible to create interaction plots specifically exploring the nature of interactions. Such plots are provided in Figure 8.12 and 8.13, and these may be thought of as representing the edges of the response surface plot shown in Figure 8.11. Figure 8.12 shows that when increasing the molar ratio, the yield of the desired product diminishes. However, the influence of the molar ratio is greater when the reaction temperature is set to its low level than to its high level. In other words, *the effect of the molar ratio depends on the level of the reaction temperature*. In a similar way, we see in Figure 8.13 that when the

reaction temperature is raised the yield is increased. This influence is slightly more pronounced when the molar ratio is high.
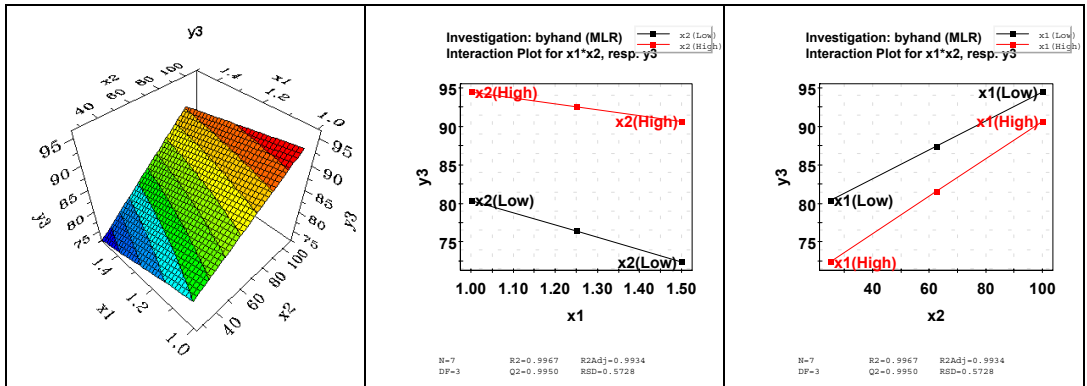
Figure 8.11: (left) Response surface plot of yield.
Figure 8.12: (middle) Interaction plot of molar ratio*reaction temperature.
Figure 8.13: (right) Interaction plot of molar ratio*reaction temperature.

Interestingly, the interaction plot may be used to uncover the strength of an interaction. Figures 8.14 to 8.16 show what the interaction plot may look like when dealing with almost no interaction (Figure 8.14), a mild interaction (Figure 8.15), and a strong interaction (Figure 8.16).
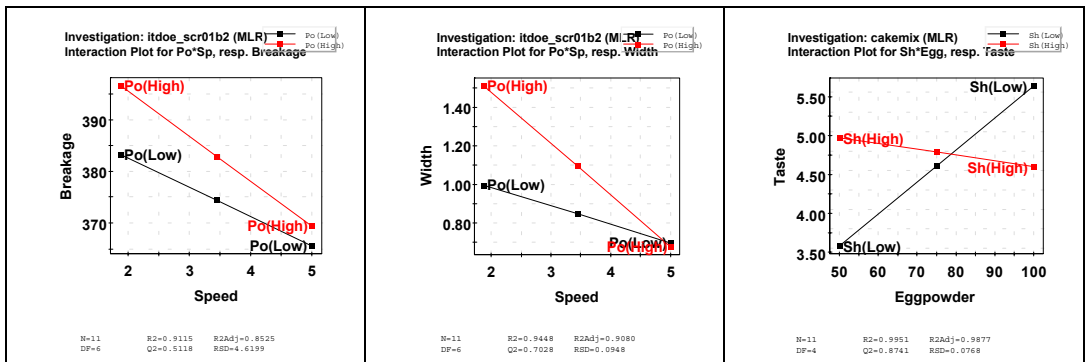
Figure 8.14: (left) Almost no interaction.
Figure 8.15: (middle) Mild interaction.
Figure 8.16: (right) Strong interaction.

Interpretation of main and interaction effects in the 2^3 case

We have now examined main and interaction effects in the 2^2 case, and seen how these may be interpreted by means of simple geometrical concepts. The geometric interpretation of main and interaction effects in a 2^3 factorial design is no more difficult. Consider the CakeMix application. Both the design matrix and the experimental matrix are shown in Figure 8.17, and the geometry of the experimental design is plotted in Figure 8.18. Notice that for clarity only the corner experiments are plotted.

	Design Matrix			Experimental Matrix			
Exp. No.	Flour	Shortn.	Egg	Flour	Shortn.	Egg	TASTE
1	-	-	-	200	50	50	3.52
2	+	-	-	400	50	50	3.66
3	-	+	-	200	100	50	4.74
4	+	+	-	400	100	50	5.2
5	-	-	+	200	50	100	5.38
6	+	-	+	400	50	100	5.9
7	-	+	+	200	100	100	4.36
8	+	+	+	400	100	100	4.86
9	0	0	0	300	75	75	4.68
10	0	0	0	300	75	75	4.73
11	0	0	0	300	75	75	4.61

Figure 8.17: Design matrix and experimental matrix of the CakeMix application.
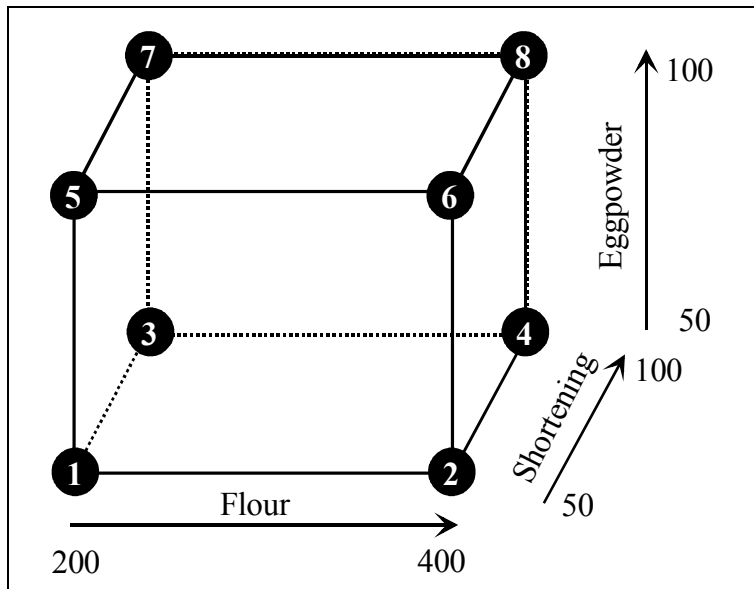
Figure 8.18: Geometry of CakeMix design.

With the experimental design displayed in Figure 8.17, three main effects, Flour, Shortening, and Eggpowder, and three interaction effects, Flour*Shortening, Flour*Eggpowder, and Shortening*Eggpowder, are determinable. Figure 8.19 shows how the main effect of flour is interpretable as the difference between the gray plane average and hatched plane average, that is, right side of cube minus left side of cube. Similarly, Figures 8.20 – 8.24 illustrate the geometric interpretation of the remaining five effects. They are all interpretable as the difference between the averages of two planes, the gray plane minus the hatched plane.
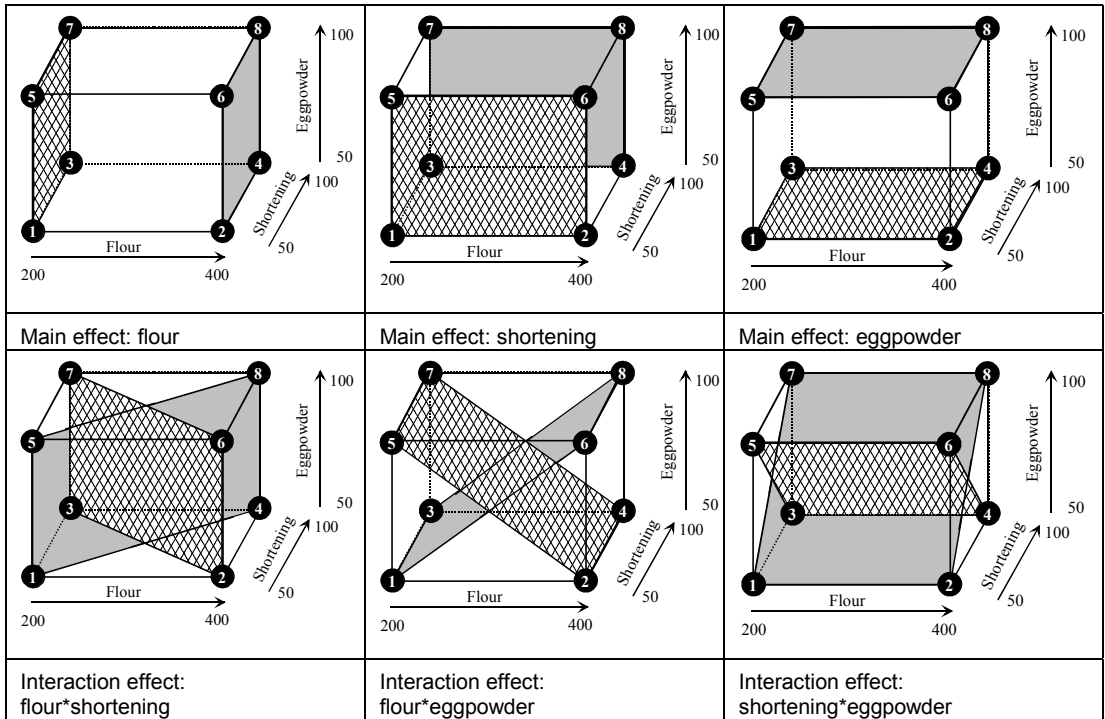
Figure 8.19: (upper left) Main effect of flour.
Figure 8.20: (upper middle) Main effect of shortening.
Figure 8.21: (upper right) Main effect of eggpowder.
Figure 8.22: (lower left) Interaction effect of flour*shortening.
Figure 8.23: (lower middle) Interaction effect of flour*eggpowder.
Figure 8.24: (lower right) Interaction effect of shortening*eggpowder.

Computation of effects using least squares fit

It would be tiresome if factor main effects and two-factor interactions had to be calculated by any of the by-hand principles outlined in this chapter. Fortunately, in reality, we may analyze DOE data by calculating a regression model using least squares fit. The rationale for the by-hand methods is that they assist in gaining an understanding of the main and interaction effects concepts. Their big disadvantage, however, is that they are sensitive to slight errors in the factor settings, such as when, for example, the reaction temperature became 27ºC instead of the wanted value 25ºC. Experimental designs with such failing data are hard to analyze with the by-hand methods just described. Least squares analysis is much less sensitive to this problem.

In fact, least squares analysis has a number of advantages, notably, (i) the tolerance of slight fluctuations in the factor settings, (ii) the ability to handle a failing corner where experiments could not be performed, (iii) the estimation of the experimental noise, and (iv) the production of a number of useful model diagnostic tools. An important consequence of least squares analysis is that the outcome is *not* main and interaction effect estimates, but a regression model consisting of *coefficients* reflecting the influence of the factors. *Such a regression coefficient has a value half of that of the corresponding effect estimate.* Figures 8.25 and 8.26 show the relationship between effects and coefficients in the ByHand

example. This relationship is discussed in more detail in Chapter 9. Also notice that the effect plot (Figure 8.26) is sorted according to the size of the effect, whereas the coefficient plot (Figure 8.25) is unsorted.
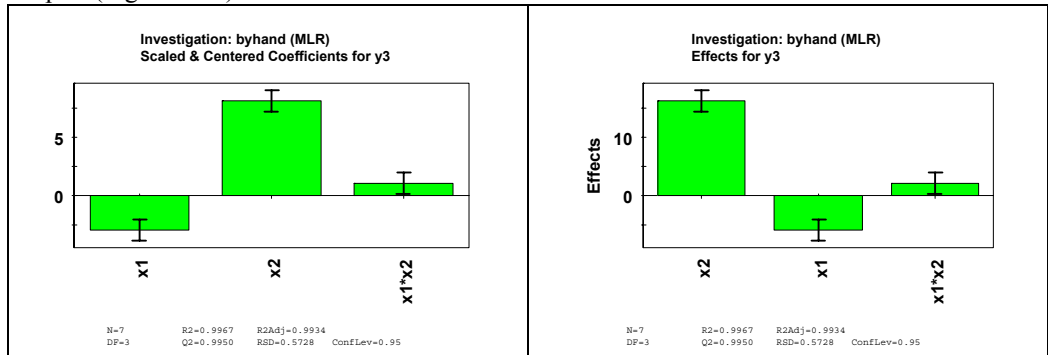
Figure 8.25: *(left) Regression coefficient plot of ByHand example.*
Figure 8.26: *(right) Effect plot of ByHand example.*

Quiz

Please answer the following questions:

How is the main effect of a factor defined?

What is the meaning of a two-factor interaction effect?

How are main and two-factor interaction effects interpretable in a 2^2 design? In a 2^3 design?

A two-factor interaction effect may be categorized in three levels of strength. Which?

Which is the quickest by-hand method for computing effects?

Give four reasons why it is better to use least squares fit for the data analysis?

What is the relationship between effects and coefficients?

Summary

In this chapter we have focused on defining, and gaining an understanding of, main and two-factor interaction effects in simple full factorial designs, such as the 2^2 and 2^3 designs. In the 2^2 case, the main effect may be geometrically understood as the difference between the average response values of two opposite edges of the square experimental region, and the interaction effect as a similar difference between the two diagonals of the square. In the 2^3 case, the main effect corresponds to the difference between the average response values of two sides of the experimental cube, and the interaction effect to a difference between two diagonal planes inserted in the three-dimensional factor space. Our reasoning here may be generalized to k factors, but this is beyond the scope of this course book.

In addition, we have introduced three methods in which simple by-hand arithmetic is used for computing main and two-factor interaction effects. The main advantage of these methods is that they give an understanding of the concepts involved. However, in reality, least squares fit of a regression model to the data is a better approach. The use of least squares fit results in regression coefficients, with half the numerical values of the corresponding effects.

9 Factorial designs (Level 3)

Objective

The aim of this chapter is to provide a deeper understanding of regression coefficients and calculated effects, and how to use their confidence intervals to reveal the most influential terms. In order to facilitate this discussion, a brief introduction to least squares analysis is given.

Introduction to least squares analysis

Consider Figure 9.1. In this graph the relationship between a single factor, called x_1, and a single response, called y_1, is plotted. Our goal is to obtain a model which can be used to predict y_1 from x_1, a common goal in, for instance, calibration studies. Because the relationship is linear the task will be to calculate the "best" straight line going through the swarm of points. However, prior to this we must first define the criterion which we are going to use for deciding when the "best" model has been found. Consider the dotted line. This line represents the best linear relationship according to a modelling criterion known as least squares. It was found with a technique known as linear regression (LR). Linear regression is based on seeking the model that minimizes the vertical deviation distances between all the experimental points and the line. An example of such a deviation, technically known as a residual, is given by the double-headed arrow. Notice that one experiment gives rise to one residual.

$$Y_1 = -1.54 + 1.61X_1 + e; \quad R^2 = 0.75$$
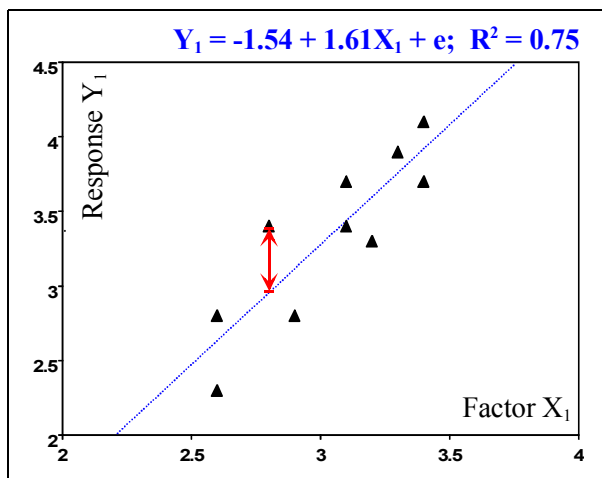
Figure 9.1: An illustration of least squares analysis.

Since some of these residuals will be positive and some negative, it is sensible to seek to minimize *the sum of the squares of the residuals*. Minimization of the residual sum of squares is a widely used criterion for finding the "best" straight line, and also explains the frequent use of the term *least squares analysis*. We can see in Figure 9.1 that the least squares line obtained has the equation $y_1 = -1.54 + 1.61x_1 + e$. The value -1.54 is called the intercept and gives the intersection of the line with the y_1-axis, that is, when $x_1 = 0$. The value 1.61 is called the gradient and indicates the slope of the line. Finally, the last term, e, represents the modelling residuals, that is, the discrepancies between the model (the line) and the reality. The R^2 parameter may be interpreted as a measure of the goodness of the model. It may vary between 0 and 1, and is a measure of how closely the line can be made to fit the points. When R^2 equals 1 the fit is perfect and all the points are situated on the line. Hence, the lower the value of R^2 the more scattered the points are. An R^2 of 0.75 indicates a rough, but stable and useful relationship, as is also evident from the figure.

Least squares analysis applied to the CakeMix data

Least squares analysis is not restricted to a one-factor-situation, but is applicable to many factors. When least squares analysis is applied to the modelling of several factors it is commonly known as *multiple* linear regression, MLR. MLR is explained in the statistical appendix. We are now going to apply MLR to the CakeMix data. Recall that the 2^3 full factorial design, augmented with three center-points, supports the interaction model $y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_{12}x_1x_2 + \beta_{13}x_1x_3 + \beta_{23}x_2x_3 + \varepsilon$. Also recall that in this equation y denotes taste, x_1 flour, x_2 shortening, and x_3 eggpowder. The model obtained was $y = 4.695 + 0.203x_1 + 0.088x_2 + 0.423x_3 + 0.038x_1x_2 + 0.053x_1x_3 - 0.602x_2x_3 + e$.

Unlike the previous example, in which a straight line was fitted to the experiments, the fitted regression model now corresponds to a twisted hyperplane in a four-dimensional space spanned by three factors and one response. We know that the hyperplane is twisted because of the large x_2x_3 two-factor interaction. This hyperplane is fitted such that the sum of the squared residuals, that is, the distances from the experimental points to the twisted hyperplane, is as small as possible. In this context, a popular way of displaying the performance of the regression model is to make a scatter plot of the relationship between measured and calculated response values. Such a plot is displayed in Figure 9.2. Apparently, the fit is excellent, because all points are located close to the 1:1 line and $R^2 = 0.99$. However, we can see that there are some small deviations between the measured and calculated response values, i.e., small *vertical* distances from the points onto the regression line. An average estimate of these deviations is given by the residual standard deviation, RSD.
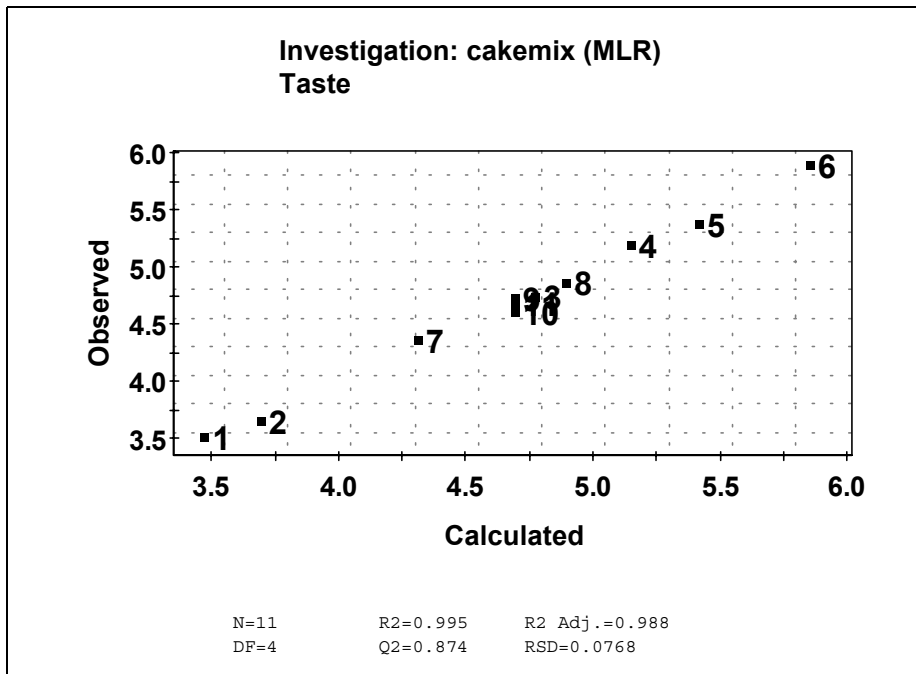
Figure 9.2: *The relationship between calculated and observed response values of the CakeMix application.*

The proper way of expressing regression coefficients

In the two examples that we have discussed so far in this chapter, the regression coefficients obtained were not listed according to the same principles. With the one factor/one response setup in the first example, the derived equation was $y_1 = -1.54 + 1.61x_1 + e$. Here the regression coefficient, 1.61, is *unscaled* and refers to the original measurement scale of the factor x_1. As a consequence, the constant term, -1.54, represents the estimated value of y_1 when x_1 is zero.

In the CakeMix case the regression model was $y = 4.695 + 0.203x_1 + 0.088x_2 + 0.423x_3 + 0.038x_1x_2 + 0.053x_1x_3 - 0.602x_2x_3 + e$. In this equation, the regression coefficients are *scaled and centered*. This means that they are no longer expressed according to the original measurement scales of the factors, but have been re-expressed to relate to the coded $-1/+1$ unit. Therefore, the constant term, 4.695, relates to the estimated taste at the design center-point, that is, when the factors have the value zero *in the coded unit*. The constant term does *not* relate to the *natural zero*, that is, zero grams of flour, shortening, and eggpowder, as this is a totally irrelevant cake mix composition.

The two kinds of regression coefficients are compared in Figures 9.3 and 9.4 using the CakeMix example. We can see that the regression coefficients differ dramatically. Figure 9.4 shows another disadvantage of using unscaled coefficients, namely that they are extremely difficult to interpret. According to the unscaled coefficients, shortening is more important than flour for improving the taste, whereas in reality the reverse is true. It is therefore common practice in DOE to express the regression results in terms of scaled and centered coefficients, as this enhances the interpretability of the model.

Figure 9.3: *(left) Scaled and centered regression coefficients of the CakeMix application.*
Figure 9.4: *(right) Unscaled regression coefficients of the CakeMix application.*

Use of coefficient and effect plots

It is convenient to display regression coefficients in a bar chart. Figure 9.5 presents the scaled and centered coefficients of the CakeMix model. On each bar, the corresponding 95% confidence interval is superimposed. We will here concentrate on how to use confidence intervals, whereas their mathematical background is dealt with in Chapter 22. Confidence intervals are error-bars which indicate the uncertainty of each coefficient. Their size depend on three factors, (i) the quality of the experimental design, (ii) the goodness of the regression model, and (iii) the number of degrees of freedom. In principle, this means that the narrowest limits are obtained with (i) a perfect design with no geometrical defects, (ii) a model with low RSD, and (iii) enough experiments.
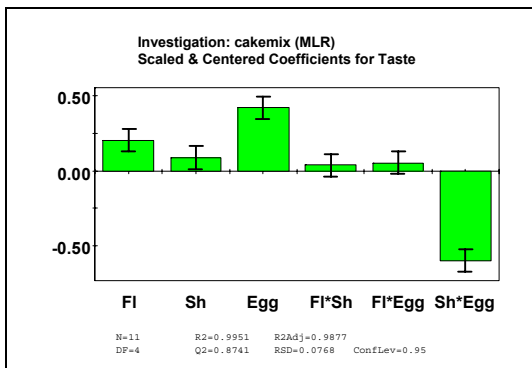
Figure 9.5: *(left) Regression coefficients of initial CakeMix model.*
Figure 9.6: *(right) Regression coefficients of refined CakeMix model.*
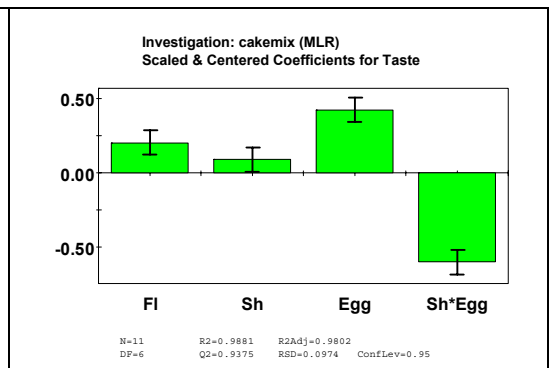
We can conclude that in the CakeMix case two coefficients, those pertaining to the two-factor interactions flour*shortening and flour*eggpowder, are statistically insignificant as their confidence intervals include zero. Hence, they can be removed and the model refitted. The results after refining the model are displayed in Figure 9.6. The numerical values of the

four remaining coefficients and their confidence intervals have not changed to any appreciable extent. We also observe that the main effect of shortening is a borderline case according to the confidence interval assessment. However, this term is allowed to stay in the pruned model because it contributes to a highly significant two-factor interaction. This is because there exists a hierarchy among model terms, and a term of lower order should not be deleted from the model if it participates in the formation of a higher order term.

Now, we turn to Figures 9.7 and 9.8, which show plots of effects instead of coefficients, prior to, and after, model refinement. Note that these effect plots are sorted according to the numerical size, and that the effect is twice as large as the coefficient. In the effect plots, the confidence intervals are displayed as two lines, and any model term that crosses these lines is considered statistically significant.
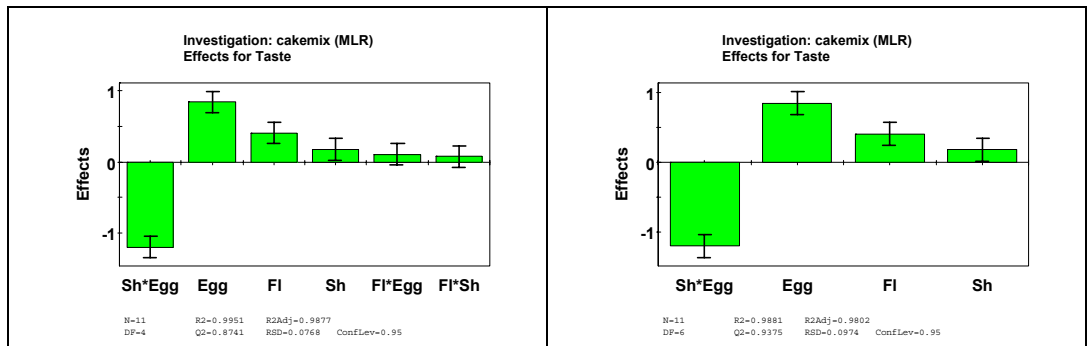
Figure 9.7: (left) Effects of initial CakeMix model.
Figure 9.8: (right) Effects of refined CakeMix model.

Other effect plots

Factor effects may also be displayed by plots other than bar charts. We will now overview some of these plotting alternatives using the refined model of the CakeMix application as an illustration. The three main effects of this model can be plotted separately, as seen in Figures 9.9 – 9.11. In comparison with the shortening*eggpowder two-factor interaction, displayed in Figure 9.12, the main effects are smaller in magnitude and of lesser importance. This is the same interpretation as before. Traditionally, a third way of plotting factor effects has been in use: the creation of normal probability plots (N-plots) of effects. But since such plots are less useful than those described here, N-plots of effects are not recommended.
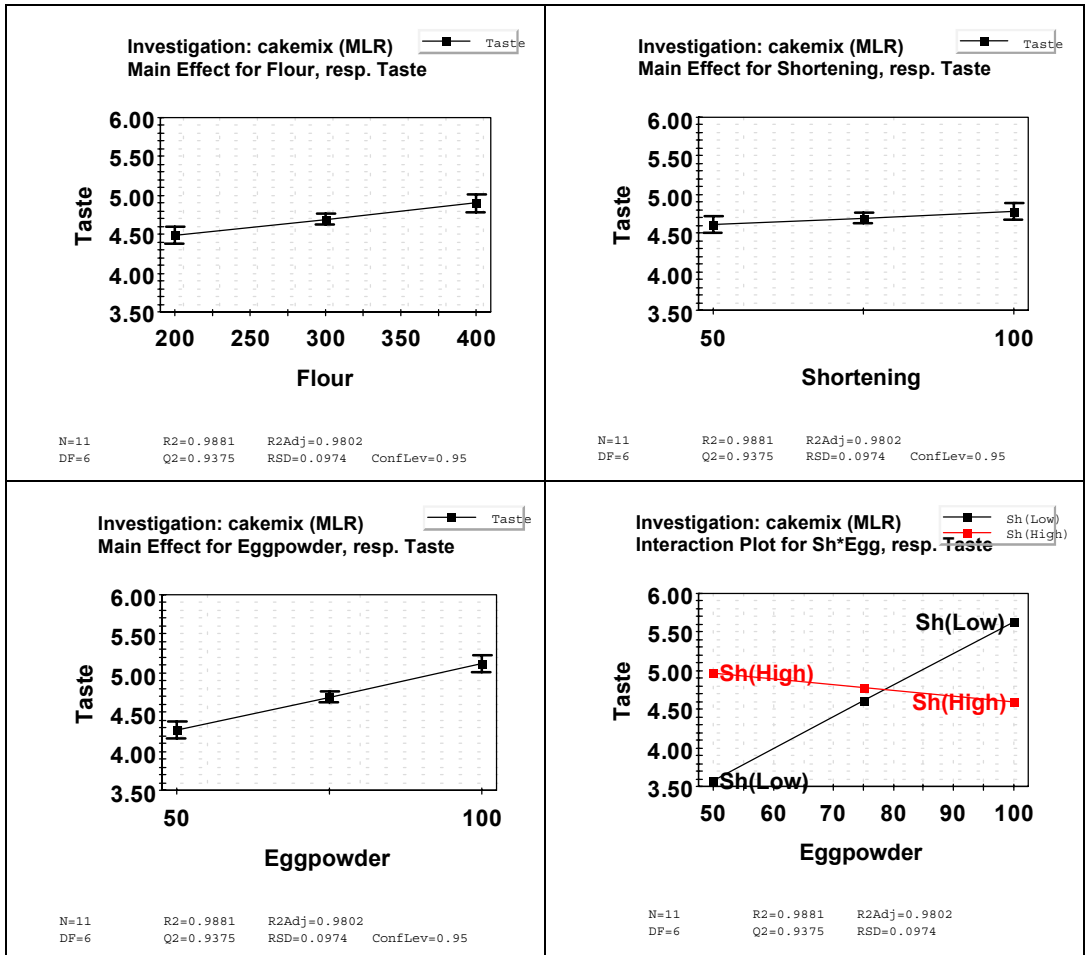
Figure 9.9: (upper left) Main effect plot of flour.
Figure 9.10: (upper right) Main effect plot of shortening.
Figure 9.11: (lower left) Main effect plot of eggpowder.
Figure 9.12: (lower right) Interaction effect plot of shortening with eggpowder.

Quiz

Please answer the following questions:

What is least squares analysis?

What is a residual?

What does R^2 signify?

What are unscaled coefficients?

What are scaled and centered coefficients?

What does the constant term represent in the two previous questions?

How can you use confidence intervals to create a better model?

Summary

Based on the principle of least squares analysis, multiple linear regression represents a very useful method for the analysis of DOE data. MLR fits a model to the data such that the sum of the squared y-residuals is minimized. The outcome is a model consisting of regression coefficients, which are utilized to interpret the influence of the factors. Such regression coefficients are usually expressed in either an unscaled, or a scaled and centered format. Using the scaled and centered representation, a regression coefficient indicates the change in the response when the factor is raised from its zero level to its high level. In this situation, the constant term expresses the estimated average response at the design center. This is the recommended option in DOE as it will facilitate model interpretation. Using the alternative, unscaled version, a regression coefficient is expressed in the original measurement unit of the factor. In this case, the constant term relates to the situation at the natural zero, and not the coded zero. Another means of model interpretation is offered through the computed effects, which are twice as large in numerical value as the coefficients.

This chapter has summarized various plotting alternatives for displaying coefficients and effects. In addition, we have described how to use confidence intervals. Confidence intervals indicate the uncertainty in the coefficients or the effects, and are useful for identifying the most important factors. The narrowest confidence intervals are obtained with (i) a perfect design with no geometrical defects, (ii) a model with low RSD, and (iii) enough experiments.

10 Analysis of factorial designs (Level 1)

Objective

The analysis of experimental data originating from full factorial designs involves three basic steps. These steps are (i) evaluation of raw data, (ii) regression analysis and model interpretation, and (iii) use of regression model. In this chapter, our aim is to give an introductory description of these three steps. We will (i) introduce a useful plot for the evaluation of raw data, (ii) introduce the powerful diagnostic tool R^2/Q^2, and (iii) demonstrate how a regression model may be used for decision making.

Introduction to the analysis of factorial designs

The analysis of experimental data generated through DOE consists of three primary stages. The first stage, *evaluation of raw data*, focuses on a general appraisal of regularities and peculiarities in the data. In most cases, important insights can be obtained even at this stage. Such insights should be used in order to enhance the subsequent regression analysis. There are a number of useful tools available, which will be described. The second stage, *regression analysis and model interpretation*, involves the actual calculation of the model linking the factors and the response(s) together, and the interpretation of this model. It is of crucial importance to derive a model with optimal predictive capability, and we will here describe the usefulness of the R^2/Q^2 diagnostic tool. As far as the model interpretation is concerned, we will mainly consider the plotting of coefficients. Finally, in the third stage, *use of regression model*, the model obtained is utilized to predict the best point at which to conduct verifying experiments or in which to anchor a subsequent design. In this respect, the analyst may use response contour plots, response surface plots and/or an interactive optimization routine.

Evaluation of raw data – Replicate plot

The replicate plot is a useful graphical tool for evaluating raw data. In such a graph, the measured values of a response are plotted against the unique number of each experiment. We will use the ByHand and CakeMix applications to illustrate this plotting principle. Replicate plots pertaining to the ByHand application are shown in Figures 10.1 – 10.3 for the three responses *side product (y_1)*, *unreacted starting material (y_2)*, and *desired product (y_3)*.
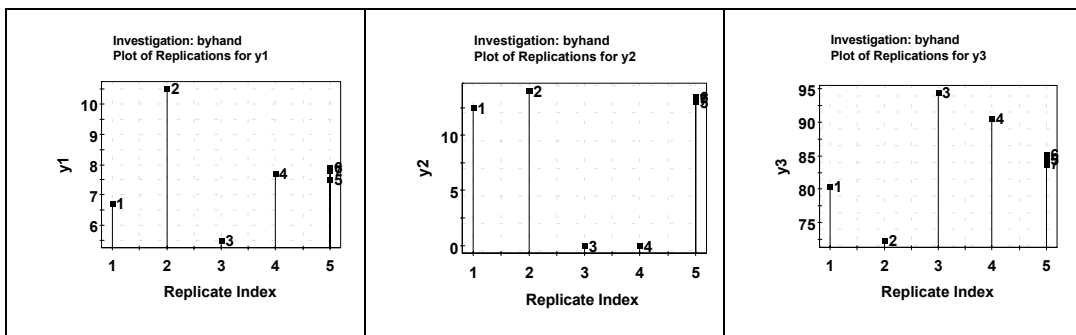
Figure 10.1: *(left) Replicate plot of side product (y_1).*
Figure 10.2: *(middle) Replicate plot of unreacted starting material (y_2).*
Figure 10.3: *(right) Replicate plot of desired product (y_3).*

In a plot of replications, any experiment with a unique combination of the factors will appear isolated on a bar, whereas experiments with identical factors settings, that is, replicates, will show up on the same bar. Hence, in the ByHand application there are three replicates with the experiment numbers 5, 6 and 7 (Figures 10.1 – 10.3). Since the variation in these three replicates is much smaller than the variation in the entire investigation series, we can conclude that the replicate error will not complicate the data analysis.

In a similar manner, Figure 10.4 represents the variation in *taste* in the CakeMix application, and apparently the replicate error is small in this case as well. Conversely, Figure 10.5 exemplifies a situation in which the replicate error is so large that a good model cannot be obtained. While not going into any details of this application, we can see that the replicates, numbered 15-20, vary almost as much as the other experiments.
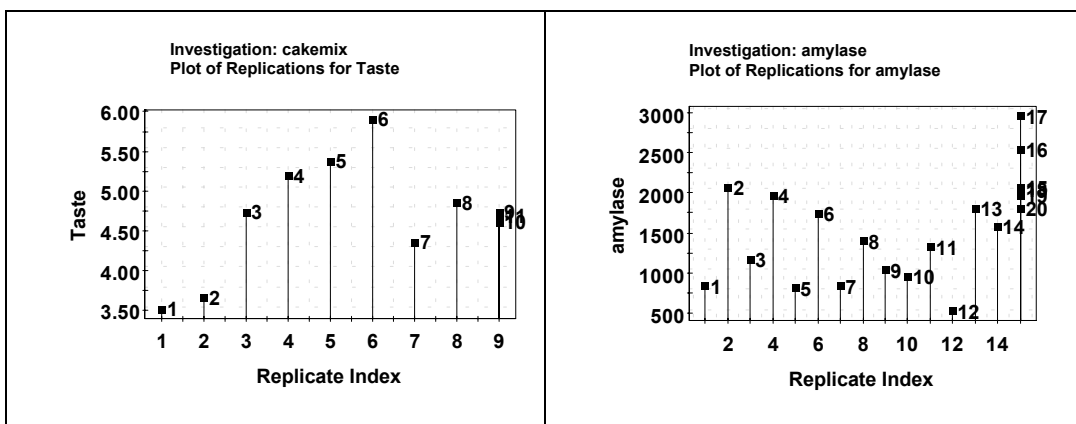
Figure 10.4: *(left) Replicate plot of taste.*
Figure 10.5: *(right) Replicate plot indicating a too large replicate error.*

Regression analysis – The R^2/Q^2 diagnostic tool

The replicate plot represents a minimum level of evaluation of raw data. Since the replicate plots did not show any problems in the ByHand and CakeMix applications, we can move on to the next stage, the regression analysis and model interpretation. When fitting a regression

model the most important diagnostic tool consists of the two companion parameters R^2 and Q^2. A plot of these for the CakeMix model is shown in Figure 10.6.
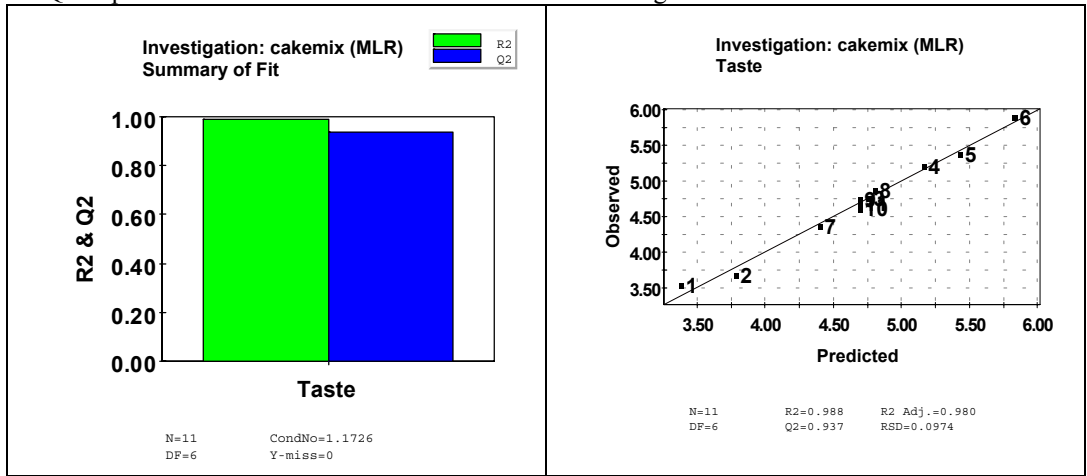
Figure 10.6: (left) Summary of fit plot for taste.
Figure 10.7. (right) Observed/Predicted plot for taste.

The left-hand bar in Figure 10.6 is R^2 and it amounts to 0.99. This parameter is called the *goodness of fit*, and is a measure of how well the regression model can be made to fit the raw data. R^2 varies between 0 and 1, where 1 indicates a perfect model and 0 no model at all. When R^2 is 1 all points are situated on the diagonal line in Figure 10.7. The main disadvantage of R^2 is that it can be made arbitrarily close to 1, by, for instance, including more terms in the model. Hence, R^2 alone is not a sufficient indicator for probing the validity of a model.

A much better indication of the validity of a regression model is given by the Q^2 parameter. Q^2 is the right-hand bar in Figure 10.6 and it equals 0.87. This parameter is called the *goodness of prediction*, and estimates the predictive power of the model. This is a more realistic and useful performance indicator, as it reflects the final goal of modelling – predictions of new experiments. Like R^2, Q^2 has the upper bound 1, but its lower limit is minus infinity. For a model to pass this diagnostic test, both R^2 and Q^2 should be high, and preferably not separated by more than $0.2 – 0.3$. A substantially larger difference constitutes a warning of an inappropriate model. Generally speaking, a $Q^2 > 0.5$ should be regarded as good, and $Q^2 > 0.9$ as excellent, but these limits are application dependent. For the relevant equations regarding how to calculate R^2 and Q^2, refer to the statistical appendix.

R^2/Q^2 pointing to a poor model

Unfortunately, the regression modelling is not always as simple and straightforward as in the previous section. We will now make a closer inspection of the regression analysis of the ByHand data, where things become a little more challenging. The R^2/Q^2 parameters are displayed in Figure 10.8. In this case, we obtain three pairs of bars, simply because the data contains three response variables.
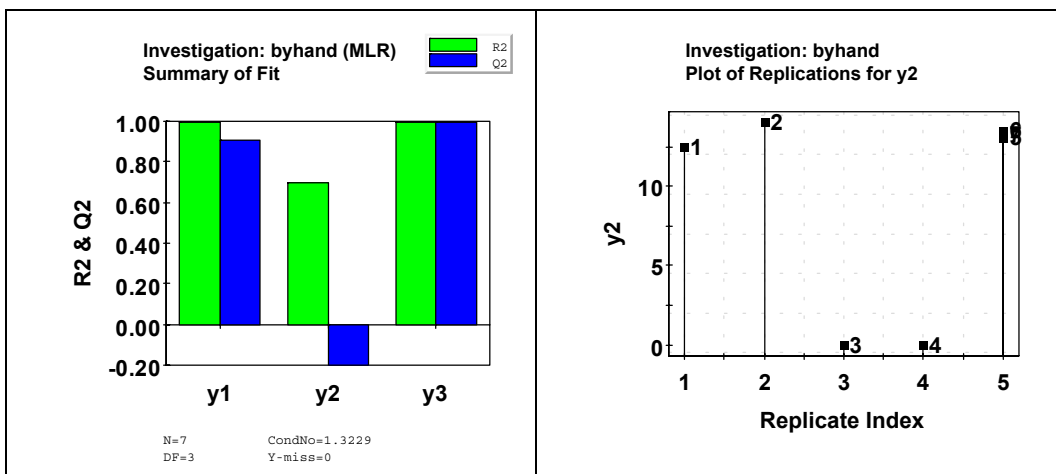
Figure 10.8: (left) Summary of fit plot of ByHand model.
Figure 10.9: (right) Replicate plot of y_2 (unreacted starting material).

According to Figure 10.8, y_1 *(side product)* and y_3 *(desired product)* are well fitted and predicted by the interaction model, since both R^2 and Q^2 are high, and closer together than $0.2 – 0.3$. In contrast, the situation is far from ideal for the second response, y_2 *(unreacted starting material)*, which has a fairly high R^2 but negative Q^2. The negative Q^2 indicates an invalid model devoid of any predictive power. Now, we have to figure out why we get this model behavior. Actually, we have already examined a plot hinting at the underlying reason - the replicate plot of y_2, which is re-displayed in Figure 10.9.

Since the response values of the replicated center-points are amongst the highest, we must have a non-linear dependence between y_2 and the two factors, molar ratio of formic acid/enamine and temperature. For a linear or an interaction model to be valid, one would expect to see the replicated center-points in the middle part of the response interval. Clearly, this is not the case and hence we can conclude that the response/factor relationship is curved. Such a curvature can only be adequately represented by a model with quadratic terms. Unfortunately, the full factorial design used here does not allow the estimation of quadratic terms. Hence, the fitted interaction model is of too low a complexity. This means that the predictive ability will break down, as manifested by the negative Q^2-value. In Chapter 12, we will see that this problem is rather easily addressed.

Model interpretation – Coefficient plot

Model interpretation also plays an important role in the data analysis. This section will describe how the coefficient plot can be used for model interpretation, and, eventually, model pruning. Consider the CakeMix data set. According to the R^2/Q^2 diagnostic tool plotted in Figure 10.10, the fitted interaction model is excellent. In this case, it is logical to interpret the model, and this is accomplished with the coefficient plot shown in Figure 10.11. According to Figure 10.11, there are two small and insignificant two-factor interactions, Fl*Sh and Fl*Egg. These terms may be omitted and the model refitted to the data.
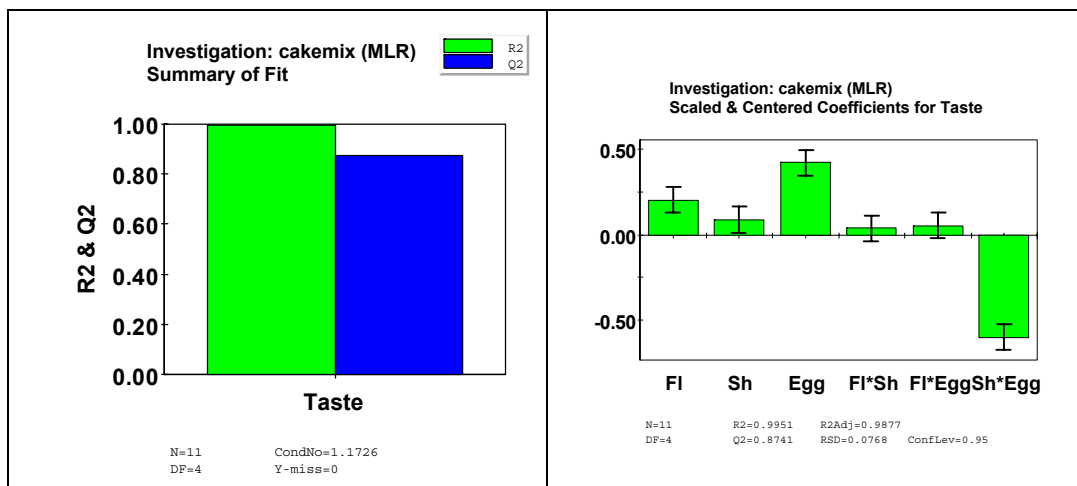
Figure 10.10: *(left) Summary of fit plot of taste.*
Figure 10.11: *(right) Regression coefficients of model for taste.*

Figures 10.12 and 10.13 show the outcome of this model refinement. Notice, that due to the near-orthogonality of the experimental design, the numerical values of the remaining coefficients have not changed. We may also observe that R^2 has undergone a tiny decrease, and that Q^2 increased from 0.87 to 0.94. The increase in Q^2 is not large, but appreciable, and indicates that we now have a simpler model with better predictive ability. The interpretation of the refined model indicates that in order to improve the taste one should concentrate on increasing the amounts of eggpowder and flour. The large two-factor interaction will be examined shortly.
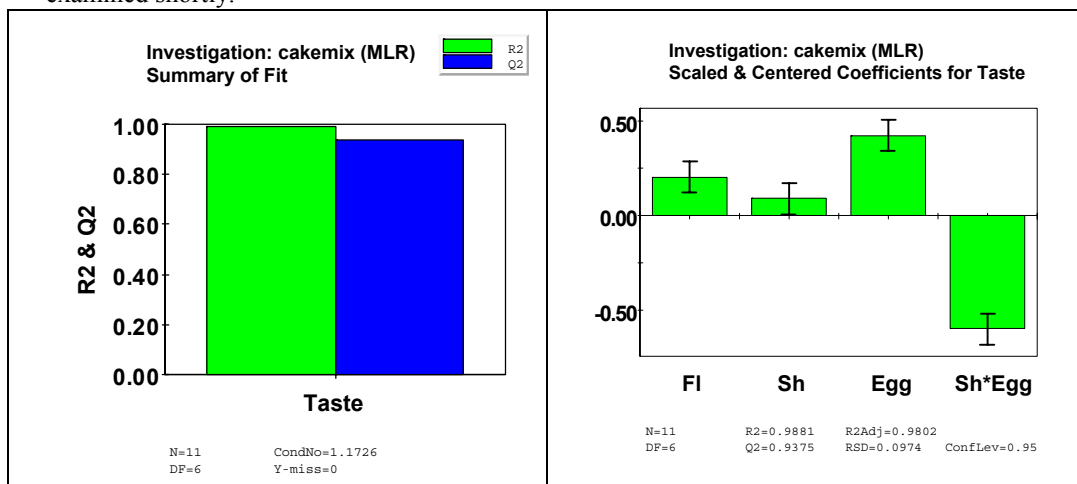
Figure 10.12: *(left) Summary of fit plot of taste, after model refinement.*
Figure 10.13: *(right) Regression coefficients of taste, after model refinement.*

Use of model – Response contour plot

When it is believed that the optimal regression model has been acquired it is pertinent to carry out the third stage of the data analysis, *use of model*. Here, the aim is to gain a better understanding of the modelled system, to decide if it is necessary to continue experimenting, and, if so, to locate good factor settings for doing this. First it must be decided which response contour plots are most meaningful. In the CakeMix data set, the strong two-factor interaction between shortening and eggpowder regulates this choice. Figure 10.14 shows a response contour plot created with the factors shortening and eggpowder as axes, and flour fixed at its high level.
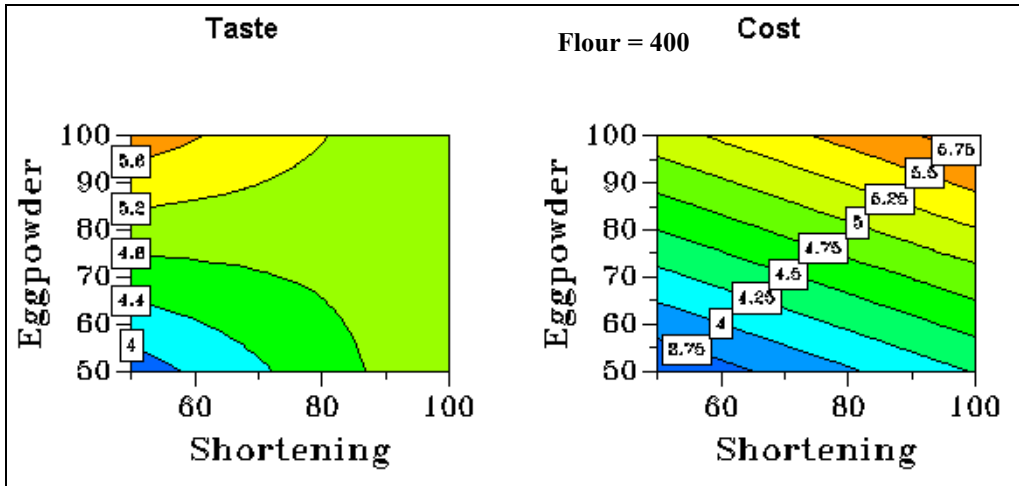
Figure 10.14: *Response contour plot of taste and cost of ingredients.*

The response contour plot in Figure 10.14 is twisted as a result of the strong two-factor interaction. Obviously, to improve the taste, we should position new (verifying) experiments in the upper left-hand corner.

In reality, however, maximizing one response variable, like taste of cake or quality of product, is usually not the single operative goal, and sometimes the most important goal is to minimize the production cost. We may transfer this reasoning to the CakeMix study. Since shortening is a cheaper ingredient than eggpowder, it may be argued that the lower right-hand corner represents a more relevant region of operability (Figure 10.14). This corner offers a reasonable compromise between high taste and low cost.

Model interpretation and use in the case of several responses

Since the CakeMix study contains only one response, the taste, only one model and hence one set of regression coefficients needs to be contemplated. This makes the model interpretation quite simple. With M responses M models are fitted with MLR, producing M sets of regression coefficients for model interpretation. Thus, there are three sets of regression coefficients to consider in the ByHand case. These regression coefficients are given in Figures 10.15 – 10.17.
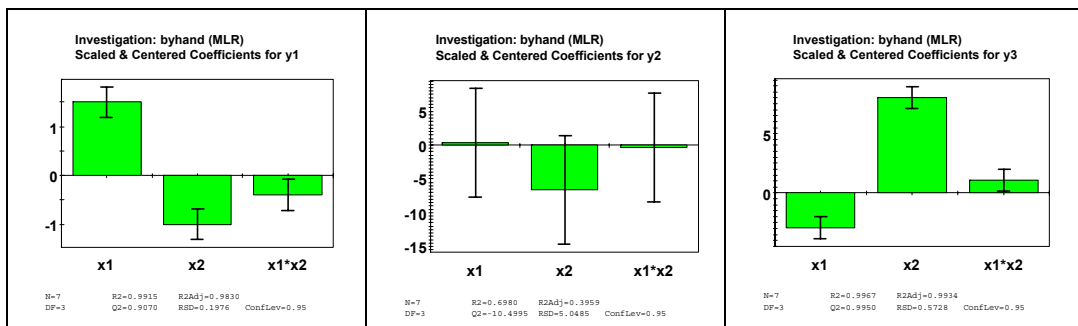
***Figure 10.15:** (left) Regression coefficients of side product (y_1).*
***Figure 10.16:** (middle) Regression coefficients of unreacted starting material (y_2).*
***Figure 10.17:** (right) Regression coefficients of desired product (y_3).*

At first glance, there are two striking features in the sets of regression coefficients (Figures 10.15 – 10.17). The most conspicuous feature is the huge confidence intervals of the model for y_2, *unreacted starting material* (Figure 10.16). But, really, this comes as no surprise, since we know that this response exhibits a quadratic behavior, which is poorly modelled by the interaction model (see Figure 10.9). The other feature is the weakness of the two-factor interaction, which is barely significant for y_1, *side product*, and y_3, *desired product*. For simplicity and clarity, however, we will keep these models.
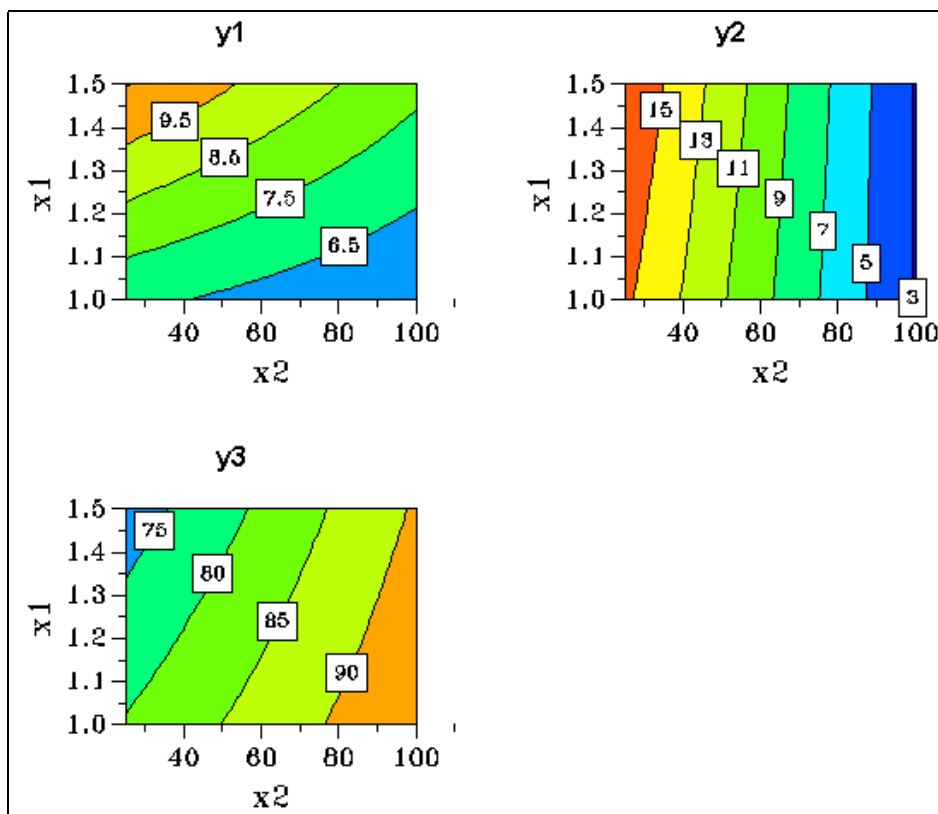
***Figure 10.18:** Triple response contour plot of side product, unreacted starting material, and desired product.*

A convenient way of surveying these models is to construct three response contour plots and place them near each other. Such a triple-contour plot is given Figure 10.18. The experimental goal was to decrease the side product, decrease the unreacted starting material, and increase the desired product. Figure 10.18 reveals that these objectives are in no conflict and are predicted to be best met in the upper left-hand corner, that is, with low molar ratio and high temperature. It is appropriate to carry out new experiments in that region. Moreover, in the interpretation of this triple-contour plot, it must be borne in mind that the model underlying the y_2 contour plot is weak, and this will result in uncertainties in the prediction phase.

Quiz

Please answer the following questions:

What are the three basic stages of data analysis?

What is the purpose of the replicate plot?

Which requirements must be met by the R^2/Q^2 diagnostic tool?

When does this diagnostic tool indicate a model of low quality?

What are the basic tools for model interpretation and model use?

Summary

The data analysis of full factorial designs can be said to comprise a sequence of three stages. The first stage is *evaluation of raw data*, in which the replicate plot is an informative tool. This plot shows the size of the replicate error in relation to the variation across the entire investigation. One example was shown, in which the replicate error was too large to permit a good model to be derived. The second stage is *regression analysis and model interpretation*. In connection with this step, we introduced a powerful diagnostic tool: the simultaneous use of the two parameters R^2 and Q^2, denoted *goodness of fit* and *goodness of prediction*, respectively. For a model to be valid, both R^2 and Q^2 must be high, as close to one as possible, and *R^2 must not exceed Q^2 by more than 0.2-0.3*. We also showed how the model in the CakeMix application was slightly improved by the removal of two small two-factor interactions, as a result of which Q^2 increased from 0.87 to 0.94. This is not a big increase, but as the model was also simplified, it well serves to illustrate the modelling message we are trying to convey, that of maximizing Q^2. In addition, attention was given to the use of the coefficient plot for model interpretation and the response contour plot for model usage. Towards the end of this chapter it was shown how triple plots of coefficients and response contour plots could be used to graphically evaluate the ByHand application and its three responses.

11 Analysis of factorial designs (Level 2)

Objective

The objective of this chapter is to deepen the understanding of the tools related to (i) evaluation of raw data, (ii) regression analysis and model interpretation, and (iii) use of regression model. In Chapter 10, the replicate plot was introduced as an efficient tool for graphical display of the response data. In this chapter, we will outline four other tools, which are useful for evaluating the raw data. These tools are called condition number, scatter plot, histogram of response, and descriptive statistics of response. Regarding the regression analysis and model interpretation, we introduced the R^2/Q^2 diagnostic tool for assessing model validity, and the regression coefficient plot for displaying the model outcome. Here, two additional model diagnostic tools will be explained. The first tool is the lack of fit test carried out as part of the analysis of variance, ANOVA, and the second tool is the normal probability plot of the response residuals. Finally, regarding the use of the regression model, a procedure for converting the information in response contour plots and response surface plots into predictions for new experiments will be exemplified.

Evaluation of raw data – Condition number

The *condition number* is a tool that can be used to evaluate the performance of an experimental design *prior to its execution*. Formally, the condition number is the ratio of the largest and the smallest singular values of the X-matrix, that is, the matrix of the factors extended with higher order terms. Informally, the condition number may be regarded as the ratio of the longest and shortest design diagonals. A schematic drawing of this is shown in Figure 11.1.
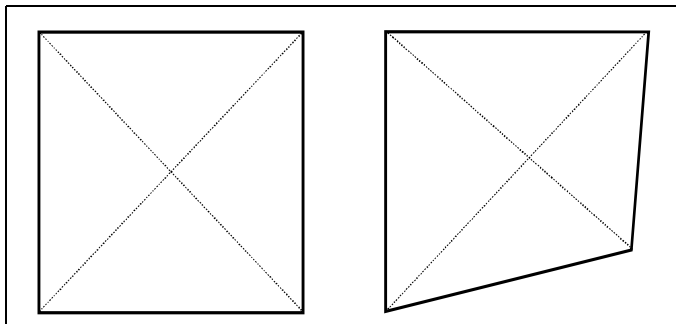
Condition Number	Objective: Screening & Robustness testing	Objective: Optimization
Good Design	<3	<8
Questionable design	3-6	8-12
BAD design	>6	>12

Figure 11.1: *A schematic illustration of the condition number. This number may be interpreted as the ratio between the longest and shortest design diagonals.*

For the left-hand design, which is symmetrical, the ratio of the two diagonals is 1. For the right-hand design, which is unsymmetrical, the ratio is > 1. From this, we see that the condition number is a measure of the *sphericity* of the design, or, its *orthogonality*. All two-level factorial designs, full and fractional, without center-points have a condition number of 1, and then all the design points are situated on the surface of a circle or a sphere.

Interestingly, by means of the condition number it is possible to formulate guidelines for the assessment of the performance of an experimental design. Such guidelines, applicable to designs in quantitative factors, are given in the lower part of Figure 11.1. With qualitative factors and mixture factors, the condition numbers are generally much higher. The listed reference values permit the evaluation of the design prior to carrying out its experiments. Hence, when a screening design has a condition number < 3 or an optimization design < 8, the designs as such are very good.

Furthermore, whenever an experimenter wishes to make a change in a design, we would advise that the condition numbers before and after the change be computed, in order to get a feeling for their magnitude. If the condition number changes by less than 0.5 we regard it as a small change, and the modification is justifiable from a theoretical point of view. If the condition number changes by more than 3.0 we regard it as a large change, and such a modification of the design would require serious reconsideration.

Evaluation of raw data – Scatter plot

The scatter plot is a useful tool in connection with the condition number assessment, particularly in investigations of limited size. When a design has been evaluated and its condition number found, one may understand its geometry by making scatter plots of the factors. We will exemplify this with the ByHand application. Figure 11.2 shows the scatter plot of the two factors *molar ratio of formic acid/enamine* and *temperature.* As seen, the design is symmetrical. It is a screening design having a condition number of 1.3. Observe that it is the existence of the three center-points that makes the condition number slightly exceed 1. Apparently, this is a good design for screening. However, were the ByHand application to be modified by excluding one corner from the experimentation, a design like that illustrated in Figure 11.3 would be obtained. This D-optimal design has a condition number of 1.6, and is therefore capable of doing a good job as a screening design, despite its skewed geometry.
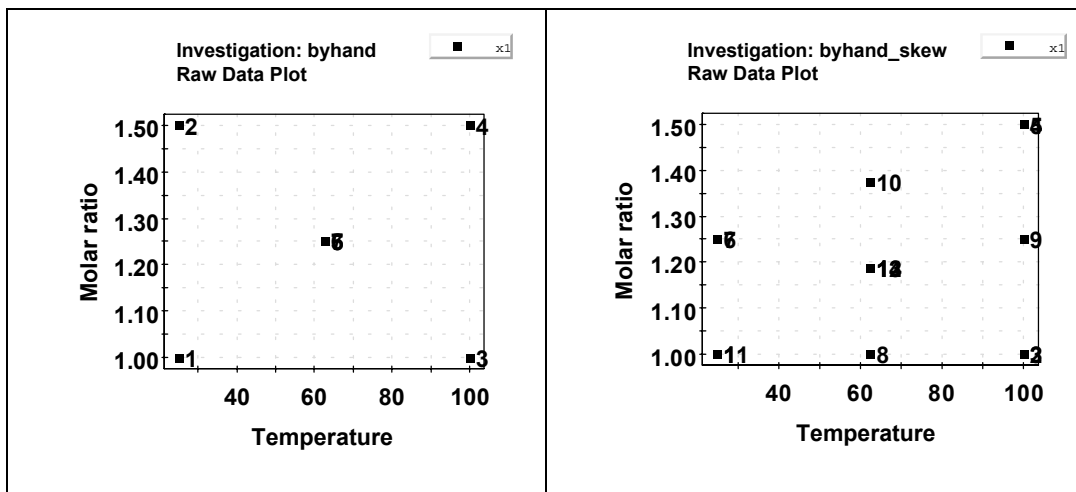
Figure 11.2: (left) Scatter plot of two factors when underlying design is of regular geometry.
Figure 11.3: (right) Scatter plot of two factors when underlying design is of irregular geometry.

The scatter plot can also be used for investigating relationships between factors and responses. In a response/factor plot we are interested in overviewing whether the relationship is linear or curved (non-linear). Figure 11.4 applies to the ByHand case and shows the relationship between the response *desired product* and the factor *temperature*. Clearly, in this case one should expect a linear relationship. In contrast, Figure 11.5 indicates a curved relationship between the response *unreacted starting material* and the factor *temperature*. Such a curved relationship will not be adequately explained by a linear or an interaction model; it needs a quadratic model. By obtaining this kind of insight prior to the regression analysis, the analyst is well prepared. Furthermore, this information also helps us to understand why we obtained such a poor model for this response (unreacted starting material) in Chapter 10. Recall that Q^2 was negative.
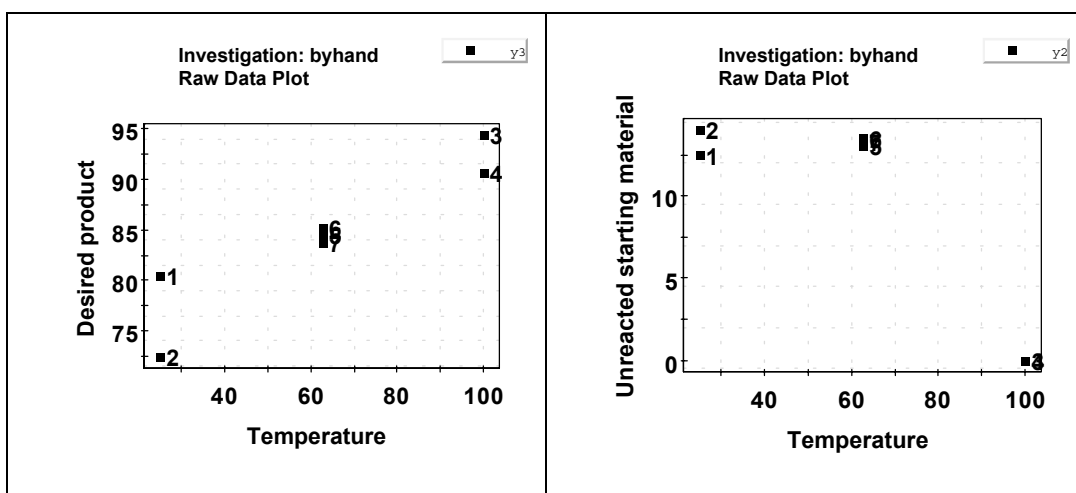
Figure 11.4: (left) Scatter plot of one response and one factor when relationship is linear.
Figure 11.5: (right) Scatter plot of one response and one factor when relationship is curved.

In summary, scatter plots of raw data make it possible to map factor/factor and response/factor relationships and gain useful insights, which can be used to guide the subsequent regression analysis. However, it must be realized that this plotting approach works best for uncomplicated investigations with few factors and responses, and is impractical with more than 4 factors and/or 4 responses.

Evaluation of raw data – Histogram of response

We will now address two tools that are useful for evaluating the statistical properties of *response* data. The first tool is *histogram of response* and the second tool is *descriptive statistics of response*. A short background to the use of these tools is warranted.

In regression analysis, it is advantageous if the data of a response variable are normally distributed, or nearly so. This improves the efficiency of the data analysis, and enhances model validity and inferential reliability. The histogram plot is useful for studying the distributional shape of a response variable. We will create some histograms to illustrate this point. Figure 11.6 shows a histogram of the response *desired product* of the ByHand application. This is a response which is approximately normally distributed, and which may be analyzed directly. As seen in the next histogram (Figure 11.7), the same statement holds true for the response *taste* of the CakeMix application.
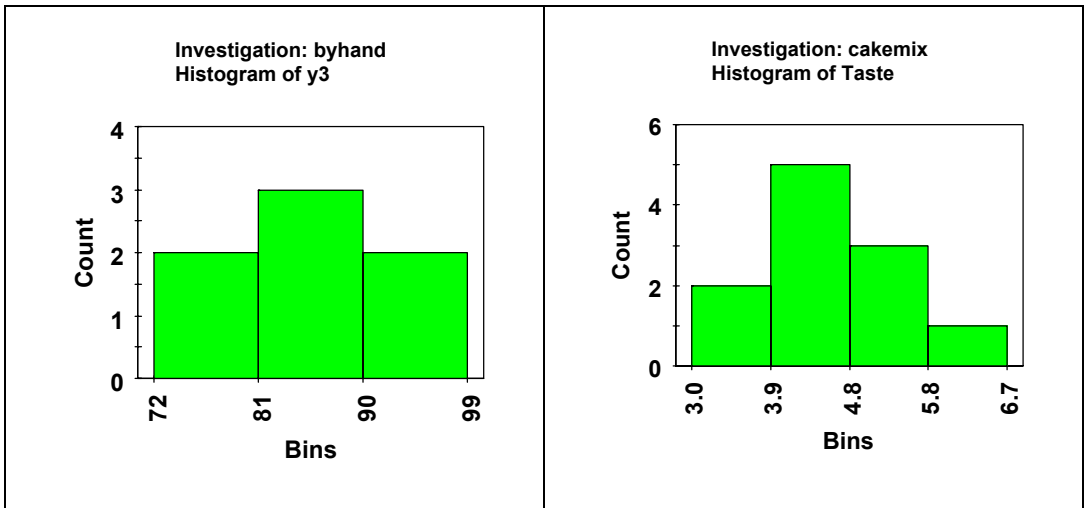
Figure 11.6: (left) Histogram of an approximately normally distributed response.
Figure 11.7: (right) Histogram of an approximately normally distributed response.

However, the third histogram (Figure 11.8), pertaining to General Example 3, shows that the response *skewness of weld* is not approximately normally distributed. This histogram has a heavy tail to the right, and indicates that one measurement is not as the others. It is much larger. It is not recommended to apply regression analysis to a response with this kind of distribution, as that would correspond to assigning the extreme measurement an undue influence in the modelling. Fortunately, it is easy to solve this problem. A simple logarithmic transformation of the response is all that is needed. Indeed, after a logarithmic transformation the extreme measurement is much closer, and hence more similar, to the majority of the data points (Figure 11.9).
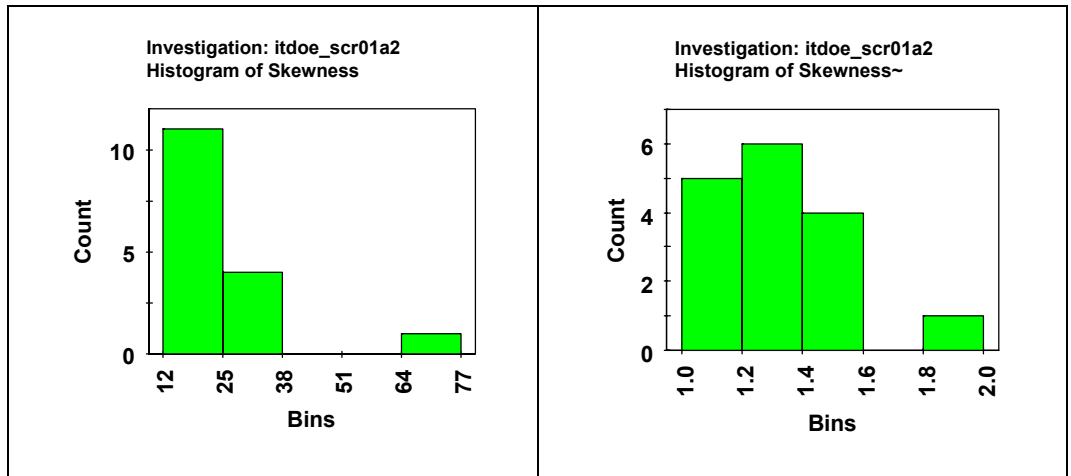
Figure 11.8: (left) Histogram with a heavy tail to the right.
Figure 11.9: (right) Histogram of a log-transformed variable.

In summary, the histogram is a powerful graphical tool for determining if a transformation of responses is needed. Also notice, that in DOE it is less common to transform a factor after the design has been executed. The reason for this is that the symmetry and balance of the design then degrade substantially (see discussion in Chapter 5).

Evaluation of raw data – Descriptive statistics of response

Another tool for investigating properties of responses is called descriptive statistics of response. It is often used in conjunction with histograms, especially when there are only a few measured values. If there are few measured values it is difficult to define appropriate data intervals for the histogram, and the histogram will then have a rugged appearance. To facilitate the understanding of the descriptive statistics tool we will compare its results with the histograms. Figures 11.10 – 11.12 show histograms of three responses which are (i) nearly normally distributed, (ii) positively skewed, and (iii) negatively skewed.

Investigation: cakemix
Descriptive Statistics for Taste

Investigation: itdoe_scr01a2
Descriptive Statistics for Skewness

Investigation: microtox
Descriptive Statistics for V11
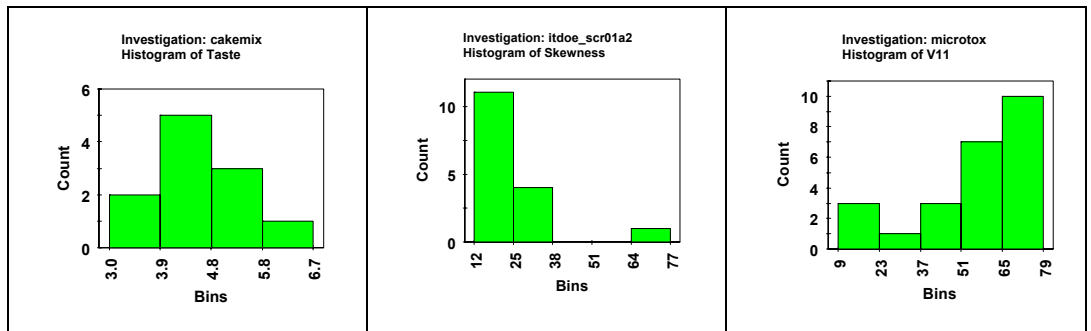
Figure 11.10: *(upper left) Histogram of a nearly normally distributed response.*
Figure 11.11: *(upper middle) Histogram of a positively skewed response.*
Figure 11.12: *(upper right) Histogram of a negatively skewed response.*
Figure 11.13: *(lower left) Descriptive statistics corresponding to Figure 11.10.*
Figure 11.14: *(lower middle) Descriptive statistics corresponding to Figure 11.11.*
Figure 11.15: *(lower right) Descriptive statistics corresponding to Figure 11.12.*

In MODDE, the descriptive statistics tool comprises a type of graph called a Box-Whisker plot (Figures 11.13 – 11.15). The Box-Whisker plot is made up of a rectangular body, the box, and two attached antennae, the whiskers. Whenever the two whiskers attached to the box are of similar length, the distribution of data is roughly normal. Figures 11.13 – 11.15 display Box-Whisker plots corresponding to the histograms of Figures 11.10 – 11.12, and we may see how the Box-Whisker plot assumes different shapes depending on the type of data distribution.

We will now give some details regarding the Box-Whisker plot. The lower and upper short horizontal lines denote the 5 and 95 percentiles of the distribution. In the box, the lowest long horizontal line depicts the lower quartile, the second line the median and the upper line the upper quartile. To summarize the use of the Box-Whisker plot one may say that when the whiskers are symmetrical, the response is approximately normally distributed. Whether to use a Box-Whisker plot or a histogram of a response much depends on personal preference.

In summary, we have now introduced four tools for evaluating raw data: condition number scatter plot, histogram of response and descriptive statistics of response. Together with the replicate plot these form a good basis for probing anomalies and errors in the input data. When the raw data are understood, the next phase is regression modelling and model interpretation.

Regression analysis – Analysis of variance (ANOVA)

In Chapter 10, we introduced the R^2/Q^2 diagnostic tool. We will now introduce two other diagnostic tools: first the analysis of variance, ANOVA, and its lack of fit test (current section), and then later the normal probability plot of residuals (next section). ANOVA is described in Chapter 23, and for the moment we will only focus on how to *use* this tool as a diagnostic test. ANOVA is concerned with estimating different types of variability in the response data, and then comparing such estimates with each other by means of F-tests. Figure 11.16 shows the tabulated output typically obtained from performing an ANOVA. In this case, the *taste* response data of the CakeMix application have been analyzed.

Taste	DF	SS	MS (variance)	F	p	SD
Total	11	247.205	22.473			
Constant	1	242.426	242.426			
Total Corrected	10	4.778	0.478			0.691
Regression	4	4.721	1.18	124.525	6.68E-06	1.086
Residual	6	0.057	0.009			0.097
Lack of Fit (Model Error)	4	0.05	0.012	3.413	0.239	0.111
Pure Error (Replicate Error)	2	0.007	3.63E-03			0.06
N = 11	Q2 =	0.9375	CondNo =	1.1726		
DF = 6	R2 =	0.9881	Y-miss =	0		
	R2Adj =	0.9802	RSD =	0.0974		
			ConfLev=	0.95		

Figure 11.16: Typical output from an ANOVA evaluation.

In ANOVA, two F-tests are made, and to evaluate these one examines the probability values, p, here displayed against a gray background. The first test assesses the significance of the regression model, and when $p < 0.05$ this test is satisfied. We can see that the CakeMix model is statistically significant as p equals 6.68E-06. The second test compares the model error and the replicate error. When a sufficiently low model error is obtained the model shows good fit to the data, that is, the model has no *lack of fit*. Hence, this latter test is known as the lack of fit test, and it is satisfied when $p > 0.05$. In the CakeMix case, p is 0.239, which is larger than the reference value, and therefore we conclude that the model has no lack of fit. The ANOVA table and its lack of fit test constitutes the second model diagnostic tool. Notice, however, that lack of fit cannot be assessed unless replicated experiments have been performed.

Regression analysis – Normal probability plot of residuals

Our third important diagnostic tool is the normal probability plot, N-plot, of response residuals. This is a good tool for finding deviating experiments, so called outliers. An example N-plot is shown in Figure 11.17. It displays the residuals of the NOx response of General Example 2. The vertical axis in this plot gives the normal probability of the distribution of the residuals. The horizontal axis corresponds to the numerical values of the residuals. However, the residuals are not expressed in the original unit of the NOx response, rather each entry has been divided by the standard deviation of the residuals. In this way, the scale of the horizontal axis is expressed as standard deviations, SDs.
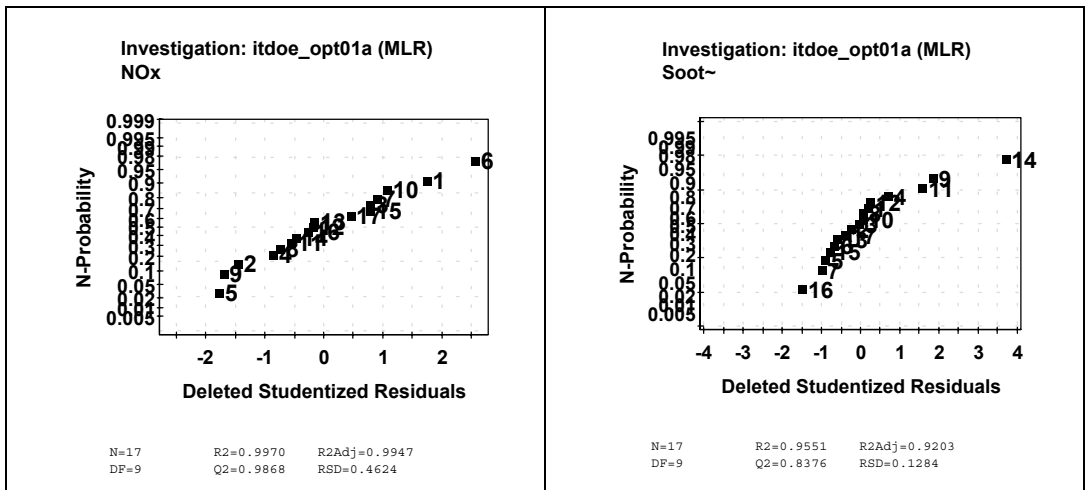
Figure 11.17: *(left) N-plot of residuals in the absence of outliers.*
Figure 11.18: *(right) N-plot of residuals when outliers are present.*

To detect possibly deviating experiments one proceeds as follows. A straight line going through the majority of the points is fitted with the eye. This line must pass through the point (0, 50%). Any point *not* falling close to this straight line is a suspect outlier. Note, that for an N-plot of residuals to be meaningful around 12-15 experiments are needed. Otherwise it is hard to draw a straight line. It is also favorable if the number of degrees of freedom exceeds 3.

In the NOx case, all the residuals are approximately normally distributed, and no deviating experiment is detectable. In the same application, however, we have another response, Soot, for which the situation is different. This is shown in Figure 11.18. In particular, experiment #14 seems to have a much larger residual than the others. This is an experiment that ought to be checked more closely.

Interestingly, it is possible to formulate warning and action limits in the N-plot. The warning limit is ± 3 SDs and the action limit is ± 4 SDs. This means that all experimental points that are found inside ± 3 SDs are good and should be kept in the model. Then there is a "gray-zone" between ± 3 and ± 4 SDs in which the analyst has to start paying attention to suspected outliers. All experiments outside ± 4 SDs are considered as statistically significant outliers and may be deleted. However, we emphasize that a critical and cautious attitude towards removing outliers from the model will be rewarding in the long run. It is always best to re-run the suspicious experiment.

In summary, we have now outlined three model diagnostic tools, the R^2/Q^2 test (Chapter 10), the lack of fit test (previous section), and the N-plot of residuals (current section). Any regression model that is to be used for predictive purposes should ideally pass these three diagnostic tools.

Use of model – Making predictions

The last stage in the analysis of DOE data is to use the regression model for making predictions. In Chapter 10, it was demonstrated how graphical tools, such as the response contour plot, could be used for this purpose. At this stage of the course, we feel that it is

---

appropriate to broaden the scope a little. Consider the response contour plot of the taste response given in Figure 11.19. This plot only indicates a *point estimate* of the taste, and does not provide the uncertainty involved. To assess the prediction uncertainty, there is the option in MODDE of transferring the most interesting factor settings into a prediction spreadsheet and computing the confidence intervals of the predicted response values.
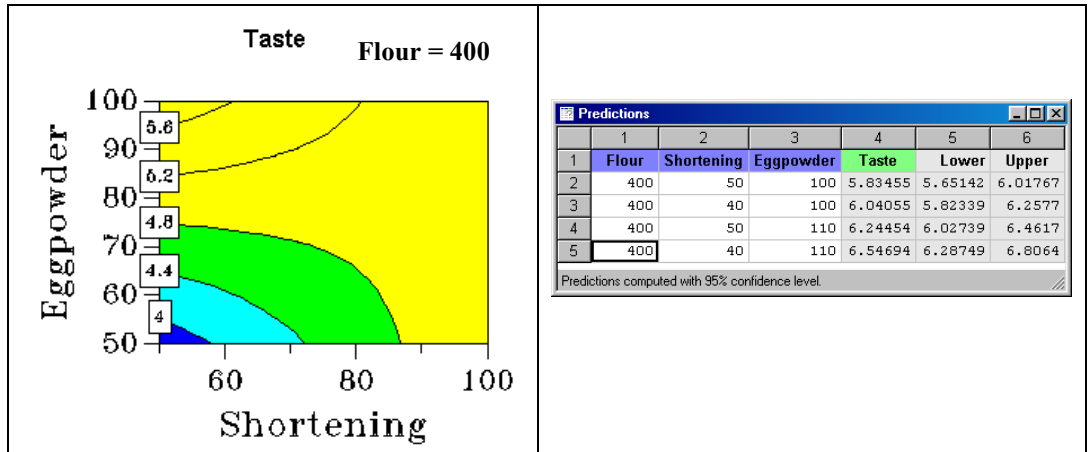
Figure 11.19: (left) Response contour plot of taste.
Figure 11.20: (right) Prediction spreadsheet providing 95% confidence intervals around predicted taste values.

Assuming that the predominant objective is that of maximizing the taste, we show a prediction spreadsheet in Figure 11.20 in which predictions are made for factor settings expected to be relevant for the goal. The first predicted point corresponds to the upper left-hand corner of the response contour plot. We can see that the 95% confidence interval indicates that at this point cakes with a taste value of 5.83 ± 0.18 are obtainable. The other three points (Figure 11.20) correspond to extrapolated factor settings outside the model calibration domain. Because they are extrapolations greater prediction uncertainties are associated with these points. However, all three predicted points unanimously indicate that even better tasting cakes are obtainable outside the explored experimental region. Hence, the process operator should select one of these proposed recipes and carry out the verifying experiment.

We note that this manual and graphically based approach works best in small designs with few runs and responses, and becomes less practical when working with many responses and sometimes conflicting goals. Fortunately, as we shall se later, it is possible to automate this kind of search for better factors settings.

Quiz

Please answer the following questions:

What does the condition number indicate?

What are the condition number limits for good screening, robustness testing and optimization designs if only quantitative factors are present?

How can the scatter plot be used in conjunction with the condition number?

What can you deduce from histograms and Box-Whisker plots of responses?

Why is it important to remove skewed data distributions prior to regression modelling?

What is analysis of variance?

What can you test with the lack of fit test?

What kind of information is retrievable from a normal probability plot of residuals?

Summary

In this chapter, we have outlined four new tools for evaluating raw data. These are condition number, scatter plot, histogram of response, and descriptive statistics of response. Together with the replicate plot, introduced in Chapter 10, these tools can help uncover anomalies in the input data. A careful evaluation of the raw data is beneficial for the subsequent regression modelling, and may result in substantially shorter analysis times. In order to facilitate the regression modelling, we also introduced two diagnostic tools, i.e., ANOVA and its lack of fit test, and the normal probability plot of residuals. With the former tool it is possible to pinpoint inadequate models, and with the latter deviating experiments. For both tools concrete reference values were specified. Together with the R^2/Q^2 diagnostic test, these two tools provide a strong foundation for assessing the validity of a regression model. Ideally, any model that is to be used for predictions should comfortably pass these three tests. Finally, a procedure for obtaining predictions of where to position new experiments was outlined. This approach utilizes the response contour and surface plots and a prediction spreadsheet for computing the prediction uncertainties in terms of 95% confidence intervals. This prediction procedure can also be used for extrapolating to new areas outside the explored experimental region.

12 Analysis of factorial designs (Level 3)

Objective

There are two main objectives of this chapter. Our first objective is to describe four common causes of poor models, describe how to pinpoint these causes, and illustrate what to do when they have been detected. Fortunately, in most cases, the measures needed to solve such problems are not sophisticated. Our second aim is to provide an introduction to the partial least squares projections to latent structures, PLS, method. PLS has certain features which are appealing in the analysis of more complicated designs, notably its ability to cope with several correlated responses in one single regression model. The PLS model represents a different mathematical construction, which has the advantage that a number of new diagnostic tools emerge. These tools are useful for model interpretation in more elaborate applications, and are more informative than other diagnostic tools.

Cause of poor model. 1. – Skew response distribution

Failure to recognize that a response variable has a skewed distribution is a common reason for poor modelling results. The need for response transformation may be detected with a histogram or a Box-Whisker plot of raw data. Consider General Example 2 and the third response called Soot. Figure 12.1 shows the histogram for Soot and Figure 12.2 the corresponding Box-Whisker plot. It is seen that this response is too skewed to be modellable, and, in fact, the modelling of this response is quite poor in comparison with the other two responses of this investigation. This is indicated in Figure 12.3 by the comparatively low Q^2 of 0.71. Recall that this is an optimization and hence rather high demands should be put on Q^2. Since the Soot response is skewed with a tail to the right the first choice of transformation is the logarithmic transformation.

Figure 12.1: *(left) Histogram of Soot.*
Figure 12.2: *(middle) Box-Whisker plot of Soot.*
Figure 12.3: *(right) Summary of fit plot for General Example 2.*

The next triplet of figures, Figures 12.4 – 12.6, show the results after log-transforming Soot. Both the histogram and the Box-Whisker plot have improved substantially. Also, the goodness of prediction parameter, Q^2, has increased from its previous value of 0.71 to 0.84, which constitutes a strong indication that the deployed transformation was sensible.

Figure 12.4: *(left) Histogram of Soot after transformation.*
Figure 12.5: *(middle) Box-Whisker plot of Soot after transformation.*
Figure 12.6: *(right) Summary of fit plot for General Example 2 after transformation.*

Often, non-normally distributed responses are skewed to the right, that is, the majority of the measured values are small except for a few cases which have very large numerical values. Typical examples of responses that adhere to this kind of distribution are variables expressed as amounts, levels, and concentrations of substances, and retention times in chromatography. Such responses share another feature: they all have a natural zero, i.e., non-negativity among the numerical values. In the case of a positively skewed distribution, the logarithmic transformation is the most common method of repairing a poor model.

In other cases one may encounter responses which are skewed to the left, that is, containing predominantly high numerical values and only a few small values. The histogram and the Box-Whisker plot are useful for detecting a negatively skewed response, and an example was given in Chapter 11. Typical examples of such responses are variables which are expressed as percentages, and where almost all measured values are close to 100% except for a few which are somewhat lower. In order to make a response with a negative skew more normally distributed, we may use a modification of the logarithmic transformation called NegLog. With this transformation each measured value is subtracted from the maximum value, and then the negative logarithm is formed. There are also many other

transformations in use, but this an advanced topic which will not be addressed further in the course.

In summary, one important reason for poor models is non-normally distributed responses. A skewed response distribution is easily detected by making a histogram or a Box-Whisker plot. However, one should observe that other diagnostic tools may also be used to reveal a response which needs to be transformed.

Benefits of response transformation

A properly selected response transformation brings a number of benefits for regression modelling. Notably it may (i) simplify the response function by linearizing a non-linear response-factor relationship, (ii) stabilize the variance of the residuals, and (iii) make the distribution of the residuals more normal, which effectively implies that outliers are eliminated. As an illustration, we shall consider a screening application aimed at producing a long-lasting device for service in aircraft construction. Ten factors were varied using a screening design consisting of 32 experiments and the measured response was the lifetime in hours. The design employed supports a linear model.

As displayed in Figure 12.7, when fitting the linear model to the data the statistics $R^2 = 0.88$ and $Q^2 = 0.71$ were obtained. These results are reasonably satisfactory. However, a closer scrutiny of the model in terms of some residual plots provides a strong warning that something is not right with the computed model. Figures 12.8 - 12.10 provide plots displaying some of the modelling problems that hopefully might be relieved with a suitable response transformation. In Figure 12.8, the observed and fitted response values are plotted against each other. From the very strong curvature seen in this plot we can conclude that the response function between the 10 factors and the lifetime response is non-linear.
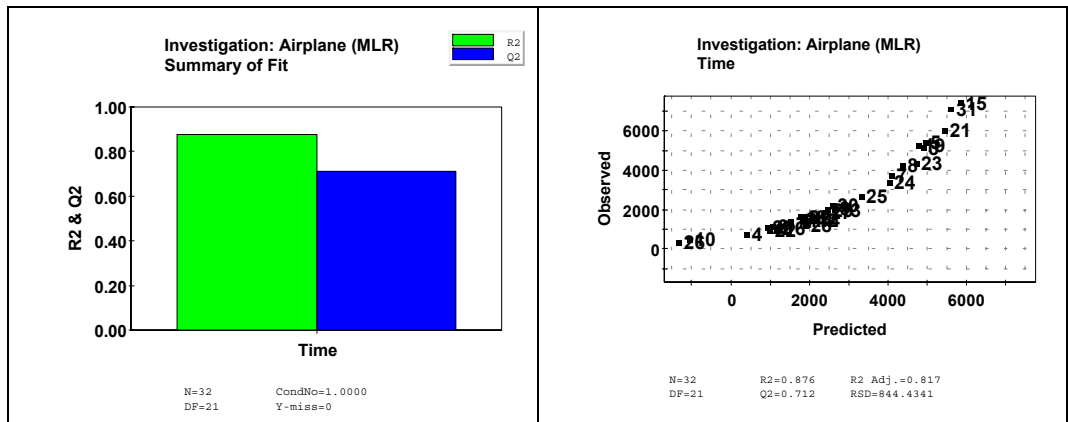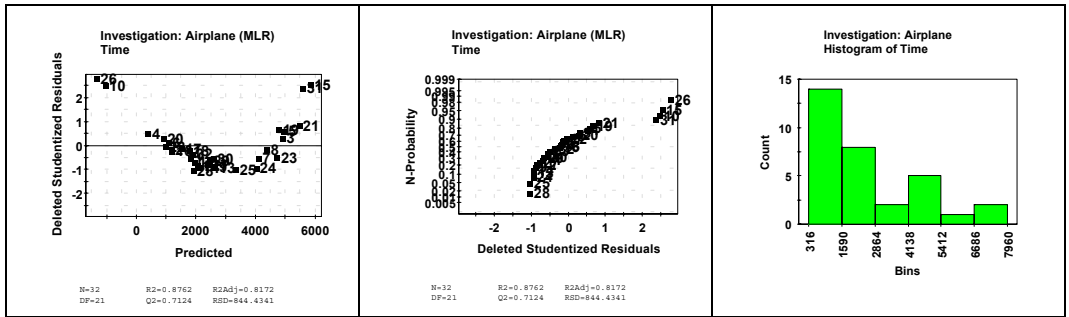
Figure 12.7: (upper left) Summary of fit plot – before transformation.
Figure 12.8: (upper right) Observed versus predicted data – before transformation.
Figure 12.9: (lower left) Residual versus predicted data – before transformation.
Figure 12.10: (lower middle) N-plot of residuals – before transformation.
Figure 12.11: (lower right) Histogram of response – before transformation.

Further, in Figure 12.9 the modelling residuals are plotted against the fitted response values. In this kind of plot one does not want to see any systematic structure. In our case, however, a strong boomerang-like data distribution is seen, suggesting some structure to the variance of the residuals. This is unwanted. Similarly, in Figure 12.10, the normal probability plot of residuals shows a group of four deviating experiments. Recall that in this plot, the desired result is that all residuals are situated on a straight line which goes through the point (0, 50%). Because of the non-linear appearance of the residuals in the N-plot, it is logical to conclude that the response is unsymmetrically distributed. That this is indeed the case is evidenced by the histogram displayed in Figure 12.11.

Consequently, the lifetime response was log-transformed and the linear model refitted, the results of which are illustrated in Figures 12.12 - 12.16. Figure 12.12 shows that R^2 now becomes 0.99 and Q^2 0.98, which are significant improvements justifying the transformation undertaken. In addition, the response transformation has (i) simplified the response function by making the response-factor relationship linear (Figure 12.13), (ii) made the size of the residual independent of the estimated response value (Figure 12.14), and (iii) made the distribution of the residuals more nearly normal (Figure 12.15). The histogram of the log-transformed response is plotted in Figure 12.16 and shows that the response is much more symmetrically distributed after the transformation.
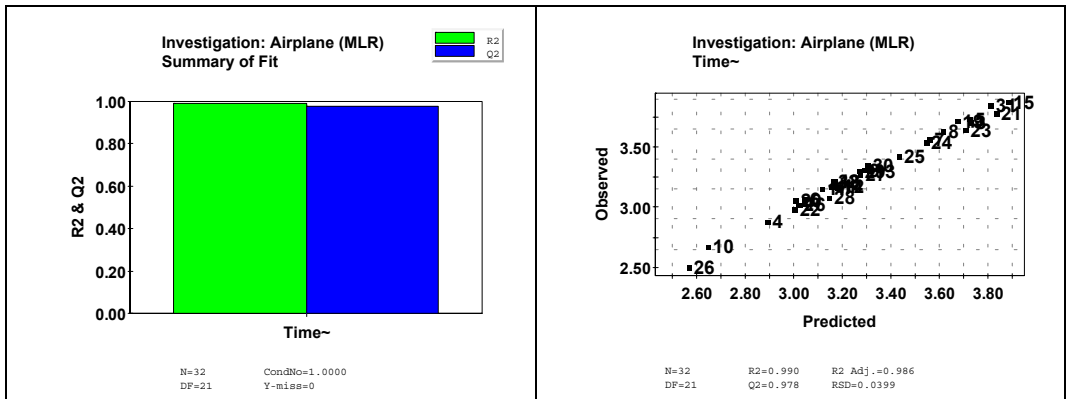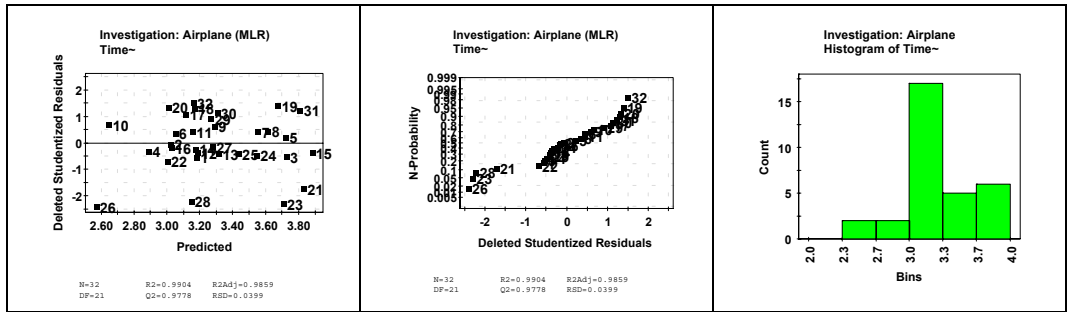
Figure 12.12: *(upper left) Summary of fit plot – after transformation.*
Figure 12.13: *(upper right) Observed versus predicted data – after transformation.*
Figure 12.14: *(lower left) Residual versus predicted data – after transformation.*
Figure 12.15: *(lower middle) N-plot of residuals – after transformation.*
Figure 12.16: *(lower right) Histogram of response - after transformation.*

In summary, the need for transforming a response is frequently found from a histogram, but may be found in many other kinds of plot, as well. And, as seen in the outlined example, a carefully selected transformation may make the regression modelling simpler and more reliable.

Cause of poor model. 2. – Curvature

Another common reason for obtaining a poor screening model is *curvature*. Curvature is a problem in screening because the normally used linear and interaction models are unable to fit such a phenomenon. Fortunately, problems related to curvature are easily detected and fixed. Let us consider the ByHand data set. We remember from Chapter 10 that a poor regression model was obtained concerning the second response, y_2, *unreacted starting material* (Figure 12.17). One clue to understanding this deficiency was found in the ANOVA table (Figure 12.18). The model for y_2 exhibits significant lack of fit, because the p-value is lower than 0.05.

y2	DF	SS	MS (variance)	F	p	SD
Total	7	883	126.143			
Constant	1	629.852	629.852			
Total Corrected	6	253.148	42.191			6.495
Regression	3	176.687	58.896	2.311	0.255	7.674
Residual	3	76.461	25.487			5.048
Lack of Fit (Model Error)	1	76.381	76.381	1909.533	5.23E-04	8.74
Pure Error (Replicate Error)	2	0.08	0.04			0.2

N = 7	Q2 = -10.4995	CondNo =	1.3229
DF = 3	R2 = 0.698	Y-miss =	0
	R2Adj = 0.3959	RSD =	5.0485
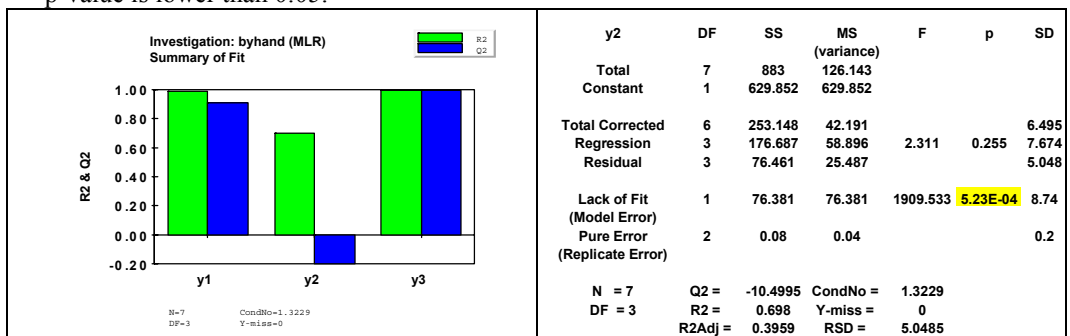
Figure 12.17: *(left) Summary of fit plot of ByHand example – interaction model.*
Figure 12.18: *(right) ANOVA table of second response – interaction model.*

In a screening application, lack of fit frequently indicates curvature. We know from previous observations that curvature exists between y_2 and the factor temperature. Lack of fit means that a regression model has some model error, that is, it contains some imperfections. Another way of formulating this deficiency in fitting ability is to state that the model error is *too large* in relation to the replicate error. It is reasonable to attempt to

modify the model so that the model error decreases. Ideally, the model error should be comparable with the replicate error.

In the ByHand case we can modify the regression model, and test whether the introduction of the squared term of temperature is beneficial. Figure 12.19 shows the summary of fitting a modified model with the term $Temp^2$, and Figure 12.20 the corresponding ANOVA table. Evidently, the introduction of a square term is necessary for the second response. However, this operation is not advantageous for the first and third responses. The new ANOVA table shows that the lack of fit has vanished, indicating that the square term is meaningful.

y2	DF	SS	MS (variance)	F	p	SD
Total	7	883	126.143			
Constant	1	629.852	629.852			
Total Corrected	6	253.148	42.191			6.495
Regression	3	252.506	84.169	393.004	2.17E-04	9.174
Residual	3	0.643	0.214			0.463
Lack of Fit (Model Error)	1	0.563	0.563	14.063	0.064	0.75
Pure Error (Replicate Error)	2	0.08	0.04			0.2

N = 7	Q2 =	0.9637	CondNo =	2.8209
DF = 3	R2 =	0.9975	Y-miss =	0
	R2Adj =	0.9949	RSD =	0.4628
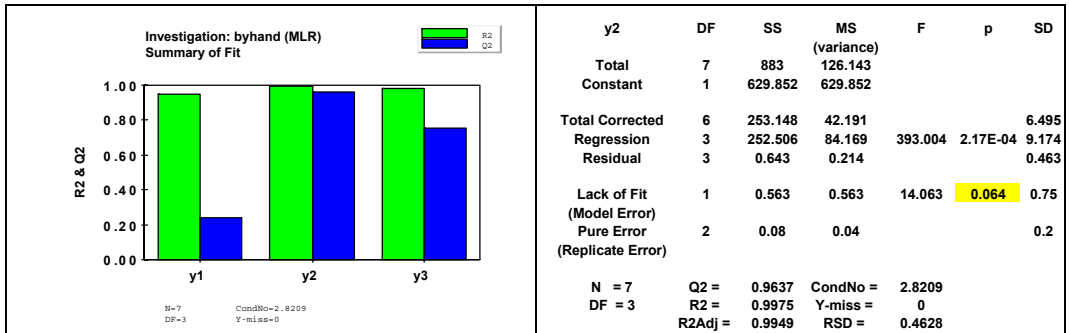
Figure 12.19: *(left) Summary of fit plot of ByHand example – modified model.*
Figure 12.20: *(right) ANOVA table of second response – modified model.*

Now, however, some words of caution are appropriate, because what we have just done is theoretically dubious. *One cannot reliably estimate a quadratic term with a screening design*. For a rigorous assessment of a quadratic term an optimization design is mandatory. Despite this, the model refinement performed suggests that the quadratic term is useful, and hence it seems motivated to expand the screening design with extra experiments, so that it becomes a valid optimization design. How this may be accomplished will be discussed later (see Chapter 13).

Cause of poor model. 3. – Bad replicates

A third common cause resulting in a poor screening model is when replicated experiments spread too much. Bad replicates are easily detected either by a replicate plot, as pointed out Chapter 10, or by studying the ANOVA table. Consider the replicate plot in Figure 12.21. It is obvious from this picture that the six replicates, numbered 15 – 20, vary almost as much as the other 14 experiments. Such a large variation between replicates will destroy any regression model.
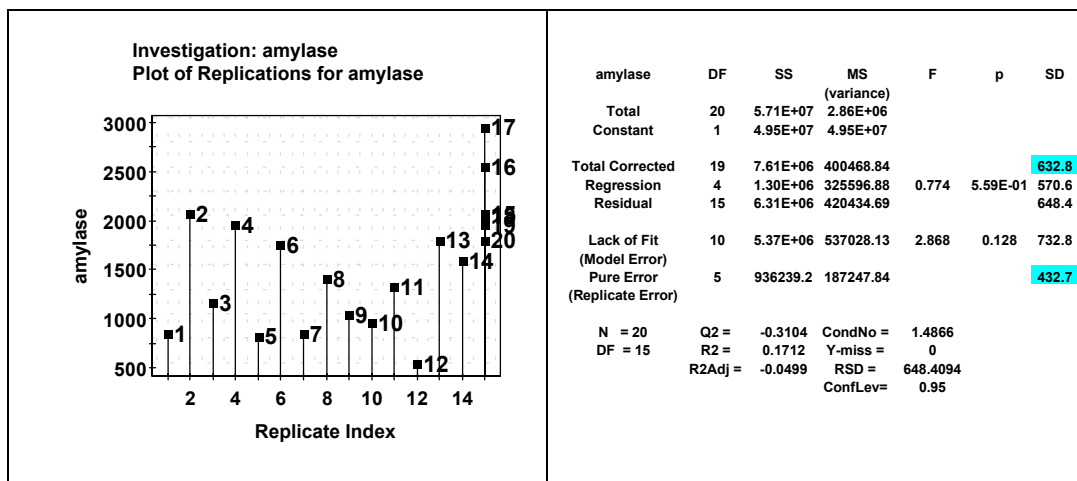
Figure 12.21: (left) Replicate plot where replicate error is too large.
Figure 12.22: (right) ANOVA table corresponding to model based on Figure 12.21.

The replicate plot gives a graphical appraisal of the relationship between the replicate error and the variation across all samples. A more quantitative assessment is retrievable through ANOVA (Figure 12.22). Focus on the two standard deviation estimates on gray background. The upper SD estimates the variation in the entire experimental design. The lower SD estimates the variation among the replicates. Clearly, these are of comparable magnitude, which is unsatisfactory. In conclusion, bad replicates are easy to detect, but harder to fix. Bad replicates means that you have a serious problem with your experimental set-up.

Cause of poor model. 4. – Deviating experiments

Deviating experiments, or *outliers*, may degrade the predictive ability and blur the interpretation of a regression model. Hence, it is of great importance to detect outliers. The normal probability plot, N-plot, of residuals, introduced and explained in Chapter 11, is an excellent graphical tool for uncovering deviating experiments. An illustration of this point is given in Figure 12.23.
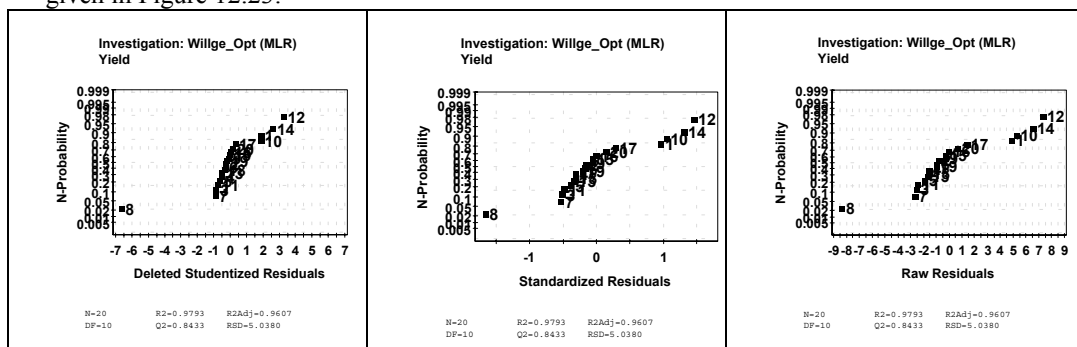
Figure 12.23: (left) N-plot of residuals – Deleted studentized residuals.
Figure 12.24: (middle) N-plot of residuals – Standardized residuals.
Figure 12.25: (right) N-plot of residuals – Raw residuals.

Without going into any details of this example, we may observe that experiment number 8 is an outlier. It is outside the action limit of ± 4 standard deviations. This deviating result could be caused by several phenomena, for instance, an incorrect experimental value in the worksheet, a failing experiment, or a too simple model. What to do with such a strong outlier is a non-trivial question. From a purely technical and statistical perspective, it is clear that the outlier should be eliminated from the model, as it degrades its qualities. However, from a scientific and ethical viewpoint, a more cautious attitude is recommended. It may well be the case that the deviating experiment is the really interesting one, and may indicate a substantially improved product or process operating condition. Automatic deletion of all outliers is not recommended, as the risk of obtaining spurious and meaningless models is increased. In the case of a failing experiment, we recommend that the initial experiment is removed from the model, but should be replaced by a re-tested run.

Moreover, the N-plot displayed in Figure 12.23 enables us to point out another important feature relating to this diagnostic test. We see that, besides experiment number 8, there are four other experiments which fall off the imaginary straight line. However, these are of lesser interest. What we are seeking in this kind of plot are the lonely and remote outliers. What happens here is that the model makes such strong efforts to model the behavior of #8 that it looses its power to explain the other four experiments.

Finally, we will draw your attention to the kind of residuals being displayed. In MODDE, there are three alternatives, called *deleted studentized residuals, standardized residuals*, and *raw residuals*. These are displayed, for the same application, in Figures 12.23 – 12.25. The raw residual is the difference between the observed and the fitted response value, expressed in the original response metric. The standardized residual is the raw residual divided by the residual standard deviation, RSD. This division makes the x-axis of the N-plot be interpretable in terms of standard deviations, but does not in any other sense change the appearance of the N-plot. The deleted studentized residual is the raw residual divided by an alternative standard deviation, computed when that particular experiment was left out of the analysis and thus not influencing the model. Deleted studentized residuals change the appearance of the N-plot in relation to the N-plot of raw residuals. One can think of the deleted studentized residual as a way of mounting an amplifier on this graphical device, simply resulting in a sharper and more efficient outlier diagnostic tool. Deleted studentized residuals require at least three degrees of freedom, and are available for MLR, but not PLS.

Introduction to PLS regression

We have now completed the treatment of four common reasons for poor models in the analysis of factorial designs. The remainder of this chapter will be devoted to the introduction of the partial least squares, PLS, regression method. A detailed account of PLS is given in Chapter 24 and in the statistical appendix. For the moment the intention is to position PLS in the DOE framework. PLS is a pertinent choice if (a) there are several correlated responses in the data set, (b) the experimental design has a high condition number, above 10, and (c) there are small amounts of missing data in the response matrix. The nice feature with PLS is that all the diagnostic tools that we have described so far are retained. In addition to R^2/Q^2, ANOVA and N-plot, PLS provides other diagnostic tools known as *scores* and *loadings*.

Consider the ByHand application. As seen in Figure 12.26, PLS experiences the same problems as MLR in the modelling of the second response, but accounts well for y_1 and y_3. The loading plot in Figure 12.27 is informative in the model interpretation. In this plot, one

can see the correlation structure between all factors and responses at the same time. Points that lie close to each other and far from the origin in the loading plot indicate variables (factors and responses) that are highly correlated. For instance, inspection of the plot in Figure 12.27 indicates that factor x_2 and response y_3 are strongly correlated. This means that x_2 has a strong positive influence on y_3 – that an increase in x_2 is likely to result in an increased value of y_3. On the other hand, factor x_1 is plotted relatively distant from this response, indeed on the opposite side of the origin – indicating that these two variables are less well correlated. We therefore expect this factor to have a lesser, and negative, influence on y_3.
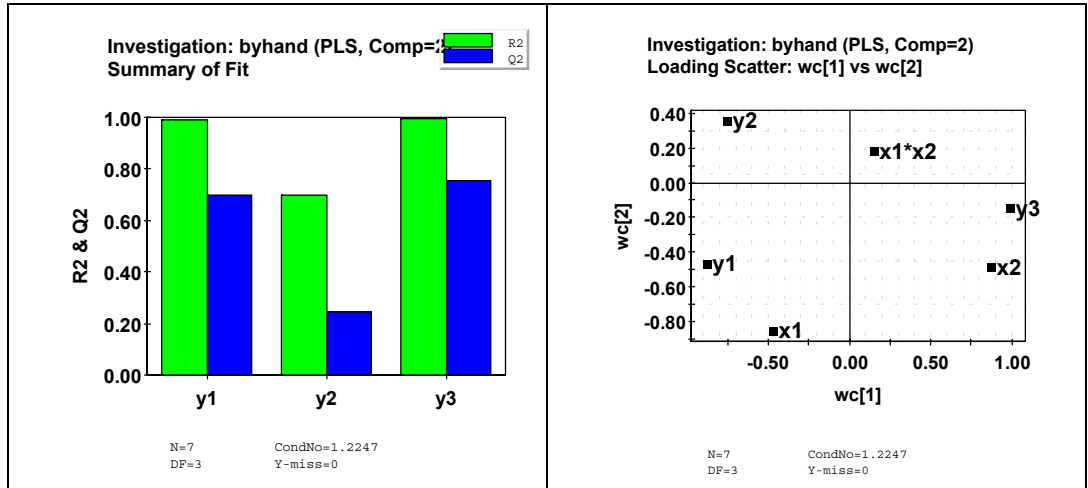
Figure 12.26: (left) Summary of fit plot of ByHand data using PLS.
Figure 12.27: (right) Loading plot of PLS model.

It is possible to convert the PLS solution into an expression based on regression coefficients. This is shown in Figure 12.28. Notice that in this plot the coefficients are expressed a little differently from usual, in that they have been normalized; that is, the raw coefficients (scaled and centered) have been divided by the standard deviation of the respective response. This gives us the best overview of the effect of each factor on each response. We can see that the information in Figure 12.28 corroborates that in Figure 12.27; namely that x_2 has a large positive coefficient for y_3, while the corresponding coefficient for x_1 is small and negative. However, these coefficient plots do not reflect the correlation structure among the responses. Fortunately, such information is evident from the loading plot (Figure 12.27), which indicates that y_1 and y_2 are positively correlated, but mutually negatively correlated with y_3. That this interpretation makes sense can be checked with the correlation matrix shown in Figure 12.29. It can be seen that the response correlation coefficients are 0.52 for y_1/y_2, -0.80 for y_1/y_3 and -0.79 for y_2/y_3. Any correlation coefficient larger than $|0.5|$ indicates a strong correlation. Hence, the three ByHand responses may well be modelled together with PLS.
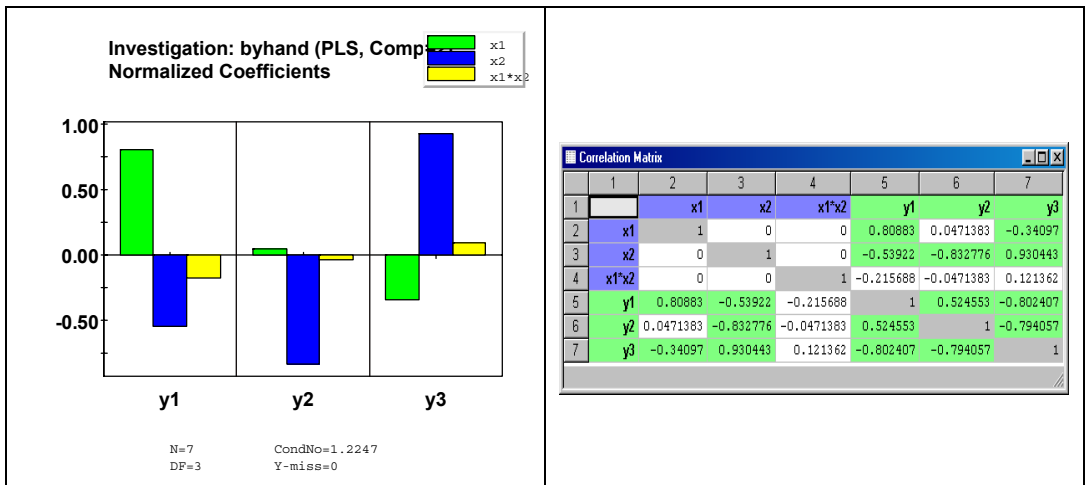
Figure 12.28: (left) Regression coefficient overview of PLS model.
Figure 12.29: (right) Correlation matrix of ByHand data.

Quiz

Please answer the following questions:

Which are the four common reasons for poor models?

What precautionary measures may be used to reduce the impact of these four causes?

Which three features of PLS make it useful for the evaluation of DOE data?

How is the PLS loading plot useful?

Summary

In the evaluation of data from full factorial designs, one can envision four common causes that may create problems in the regression modelling. These causes relate to (i) skew response distribution, (ii) curvature, (iii) bad replicates, and (iv) deviating experiments. We have described pertinent tools for detecting such troublesome features, and also indicated what to do when such problems are encountered. Towards the end of this chapter, we also introduced the PLS regression method. PLS is an attractive choice when one is working with (i) several correlated responses, (ii) a design with high condition number, or (iii) a response matrix with a moderate amount of missing data. Because PLS fits only one model to all responses, the model interpretation in terms of loadings is powerful. In addition, the interpretation of loadings may be supplemented with other tools with which we are more familiar, such as, a coefficient overview plot, a table of the correlation matrix, and scatter plots of raw data.

Part II

Chapters 13-17

The 3 General Examples

13 Experimental objective: Screening (Level 2)

Objective

Chapters 1-12 in Part I have outlined a general framework for DOE, by describing what are the key steps in the generation of an experimental design and the analysis of the resulting data. The objective of this chapter is to make use of this framework in the presentation and discussion of a DOE screening application. We will describe in detail an application dealing with the laser welding of a plate heat exchanger. In addition, we will introduce the family of fractional factorial designs, as these are the designs used most frequently in screening. Our objective is to provide a comprehensive overview of what are fractional factorial designs, and highlight their advantages and limitations. Further, the analysis of the welding data also involves the repeated use of a number of evaluation and regression analysis tools, which were outlined in the foregoing chapters. This chapter will serve as a review of the framework of DOE presented earlier. It will provide a comprehensive introduction to fractional factorial designs, and describe in detail the laser welding study. Of necessity, there will some switching between the repetition of previous concepts and their application, and new theory related to fractional factorial designs.

Background to General Example 1

Some years ago, Alfa Laval introduced a new type of plate heat exchanger, based on the principle of replacing polymer gaskets with an all-welded gasket-free plate design. The all-welded design enabled the apparatus to endure higher operating temperatures and pressures. Such plate heat exchangers are used in a wide variety of applications, for instance, in offshore oil production for cooling of hydrocarbon gas, in power plants for preheating of feed water, and in petrochemical industries for solvent recovery, reactor temperature control, and steam applications. A small cross-section of such a plate heat exchanger is shown in Figure 13.1.
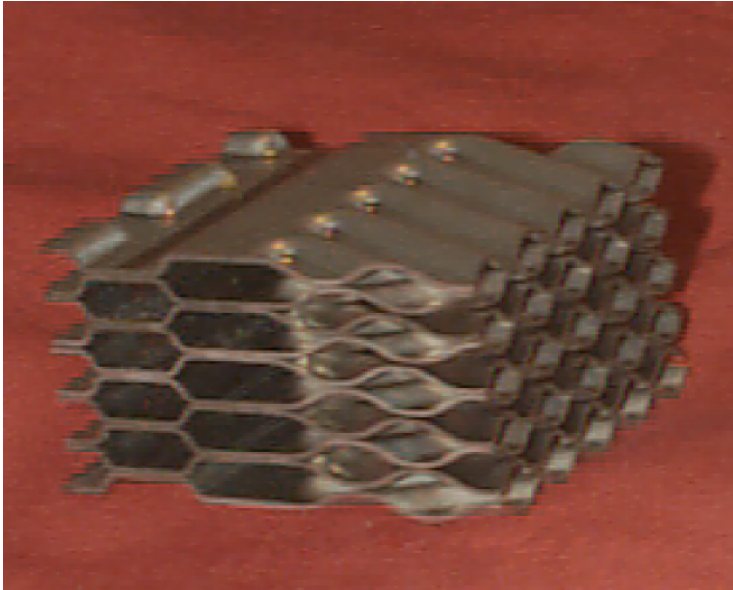
Figure 13.1: *Cross-section of plate heat exchanger.*

The example that we will present here relates to one step in the process of fabricating a plate heat exchanger, a laser welding step involving the metal nickel. The investigator, Erik Vännman, studied the influence of four factors on the shape and quality of the resulting weld. These factors were *power of laser, speed of laser*, *gas flow at nozzle* of welding equipment, and *gas flow at root*, that is, the underside of welding equipment. The units and settings of low and high levels of these factors are found in Figure 13.2. In this example, the weld takes the shape of an hour-glass, where the width of the "waist" is one important response variable. To characterize the shape and the quality of the weld, the following three responses were monitored: (i) breakage of weld, (ii) width of weld, and (iii) skewness of weld. These are displayed in Figure 13.3.

	Name	Abbr.	Units	Type	Use	Settings
1	**Power**	Po	kW	Quantitative	Controlled	2.15 to 4.15
2	**Speed**	Sp	m/min	Quantitative	Controlled	1.88 to 5
3	**NozzleGas**	No	l/min	Quantitative	Controlled	27 to 36
4	**RootGas**	Ro	l/min	Quantitative	Controlled	27 to 42

	Name	Abbr.	Units
1	**Breakage**	Br	MPa
2	**Width**	Wi	mm
3	**Skewness**	Sk	-

Figure 13.2: *(left) Factor overview of laser welding example.*
Figure 13.3: *(right) Response overview of laser welding example.*

Besides understanding which factors influence the welding process, the main goal was to obtain a robust weld (high value of breakage), of a well-defined width and low skewness. The desired response profile was a high value of breakage with target set at 385 MPa, a width of weld in the range 0.7-1.0 mm, and a low value of skewness with 20 specified as target. In the first stage, the investigator carried out eleven experiments using a medium-resolution two-level fractional factorial design. However, as we shall see during the analysis of this data-set, it was necessary to upgrade the initial screening design with more

experiments. Indeed, the experimenter conducted a further set of eleven experiments, selected to complement the first series. Towards the end of this chapter we shall analyze both sets of 11 experiments together.

Problem Formulation (PF)

Selection of experimental objective

The problem formulation (PF) process is important in DOE, and it was discussed in Chapter 4. We will now re-iterate the steps of the problem formulation by reviewing the laser welding application. The first step in the problem formulation corresponds to choosing which experimental objective to use. Recall that the experimental objective may be selected from six stages of DOE, viz., (i) familiarization, (ii) screening, (iii) finding the optimal region, (iv) optimization, (v) robustness testing, and (vi) mechanistic modelling. In the laser welding application, the selected experimental objective was screening. With screening we want to find out a little about many factors, that is, which factors dominate and what are their optimal ranges? Typically, screening designs involve the study of between 4 and 10 factors, but applications with as many as 12-15 screened factors are not uncommon. In the laser welding case there are four factors and this facilitates the overview of the results.

Specification of factors

The one thing that an experimenter always wants to avoid is to complete an experimental design, and then suddenly to be struck by the feeling that something is wrong, that some crucial factor has been left out of the investigation. If we forget an important factor in the initial design it takes a large number of extra experiments to map this factor. Thus, in principle, with DOE one can only *eliminate* factors from an investigation. In this regard, the Ishikawa, or fishbone, system diagram is a very helpful method to overview all factors *possibly* influencing the results. An example diagram is shown in Figure 13.4.
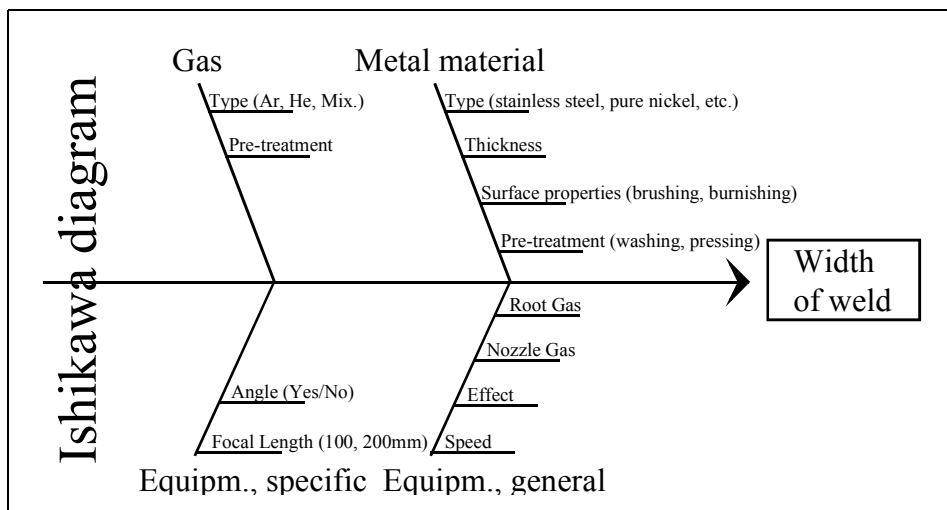
Figure 13.4: Ishikawa system diagram.

In the Ishikawa diagram a baseline is drawn. To this line is attached a number of new intersecting sub-lines corresponding to each main category of factors. Each such sub-line may be further decomposed with additional sub-lines. In the welding application we can discern four major types of factors, factors related to metal materials, factors related to weld gas, general equipment factors, and equipment factors of a more specific character. Now, let us focus on the category of metal material factors. Factors that are important in this group are the type of metal material, the thickness, the surface properties and pre-treatment procedures. The other categories of factors may be worked out in a similar fashion. The final system diagram resulting from this kind of "mental" screening of factors is helpful for overviewing the interesting factors. Correctly used, it will diminish the risk of neglecting important factors. When all factors have been listed their ranges must be defined. This is accomplished by determining the low and high investigation values for each factor. Usually, comparatively large ranges are investigated in screening, because one does not want to run the risk of overlooking a meaningful factor. The factors selected in General Example 1 were *power of laser, speed of laser*, *gas flow at nozzle*, and *gas flow at root*. Their units and settings of low and high levels are found in Figures 13.5 – 13.8. As seen, these factors are specified according to their respective untransformed metric.

Figure 13.5: (upper left) Factor definition of power of laser.
Figure 13.6: (upper right) Factor definition of speed of laser.
Figure 13.7: (lower left) Factor definition of nozzle gas.
Figure 13.8: (lower right) Factor definition of root gas.
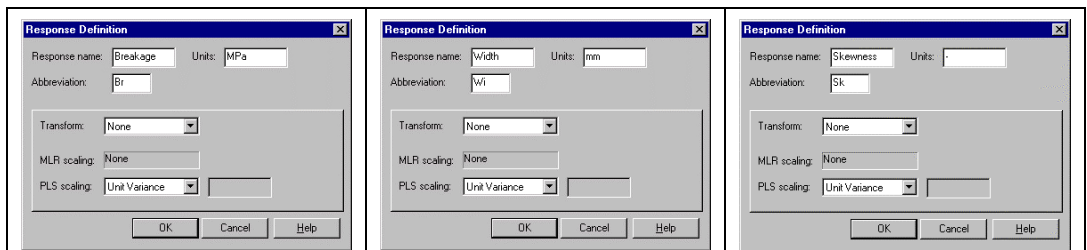
Specification of responses

The next step in the problem formulation is to specify the responses. It is important to select responses that are relevant to the experimental goals. In the laser welding application, the quality and the shape of the resulting weld were the paramount properties. Hence, the investigator registered three responses, the breakage of the weld, the width of the weld, and the skewness of the weld. It is logical to choose three responses, because product quality is typically a multivariate property, and many responses are often needed to get a good appreciation of quality. Also, with modern regression analysis tools, it is no problem to handle many responses at the same time.

Furthermore, when specifying the responses one must first decide whether a response is of a quantitative or qualitative nature. Quantitative responses are preferable to qualitative ones, because the interpretation of the resulting regression model is simplified. We see in Figures 13.9 – 13.11 that the three weld responses are quantitative and untransformed. The first response is expressed in MPa, the second in mm, whereas the third is dimensionless. The overall goal was a response profile with a high value of breakage with the target set at 385 MPa, a width of the weld in the range 0.7-1.0 mm, and a low value of skewness with 20 specified as the target.

Figure 13.9: (left) Response definition of breakage of weld.
Figure 13.10: (middle) Response definition of width of weld.
Figure 13.11: (right) Response definition of skewness of weld.

Selection of regression model

The selection of an appropriate regression model is the next step in the problem formulation. Recall, that we distinguish between three main types of polynomial models, that is, linear, interaction and quadratic models. In screening, either linear or interaction models are used. This means that in the laser welding application, one could use either model, and the final choice must depend on the expected clarity in the information and on the number of experiments allowed. With four factors, as in the laser welding study, the 2^4 full factorial design in 16 experiments is a conceivable experimental protocol, that is, a design in which all possible corners of the four-dimensional hypercube are investigated. This design supports an interaction model. To these 16 corner experiments it is recommended that between 3 and 5 replicated center-points are added, making a total of 19 – 21 experiments.
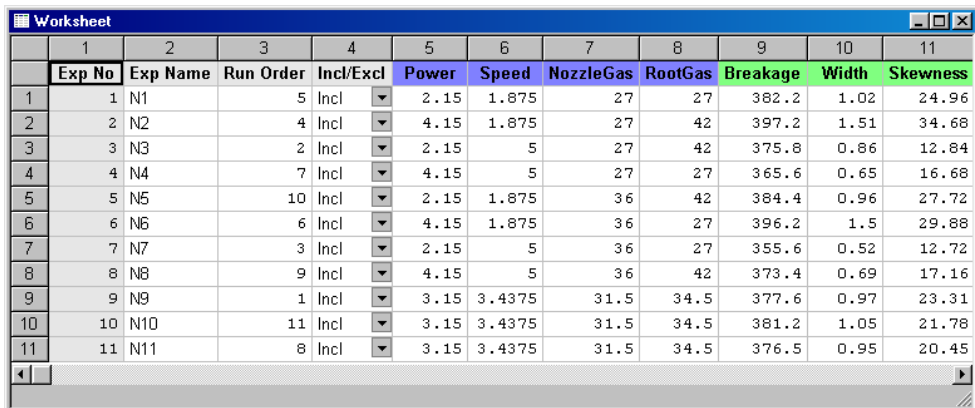
Initially, 19 – 21 runs was judged to be too many experiments. Hence, the investigator decided to restrict himself to the use of a linear model. Such a linear model is estimable with a reduced factorial design, in which only a *fraction* of all possible corners are investigated. Therefore, this type of design is called a fractional factorial design. It is depicted by the notation 2^{4-1}, which is read as a two-level experimental design in four factors, but reduced

one step. We will shortly give more details pertaining to the family of fractional factorial designs. For now, suffice it to say that the 2^{4-1} design encodes eight experiments, that is, eight corners out of the 16 theoretically possible. To these 8 experiments, the investigator appended 3 replicated center-points.

Generation of design and creation of worksheet

The last two stages of the problem formulation deal with the generation of the statistical experimental design and the creation of the associated worksheet. We will treat these two steps together in describing the laser welding application. We must realize that the chosen regression model and the design to be generated are intimately linked. Since the experimenter selected a linear model, the 2^{4-1} fractional factorial design is an excellent choice of design. This is a standard design, which prescribes eight corner experiments.

In Figure 13.12, the four worksheet columns with numbers between 5 and 8 represent the identified design. The first eight rows are the corner experiments and the last three rows are the replicated center-points. Before we leave this part of the problem formulation, we will consider the column entitled RunOrder. It contains a proposed randomized order in which to run the 11 experiments. It is recommended that all experiments be run in a randomized order, to prevent any systematic time trend from influencing the experimental values.

	1	2	3	4	5	6	7	8	9	10	11
	Exp No	Exp Name	Run Order	Incl/Excl	Power	Speed	NozzleGas	RootGas	Breakage	Width	Skewness
1	1	N1	5	Incl	2.15	1.875	27	27	382.2	1.02	24.96
2	2	N2	4	Incl	4.15	1.875	27	42	397.2	1.51	34.68
3	3	N3	2	Incl	2.15	5	27	42	375.8	0.86	12.84
4	4	N4	7	Incl	4.15	5	27	27	365.6	0.65	16.68
5	5	N5	10	Incl	2.15	1.875	36	42	384.4	0.96	27.72
6	6	N6	6	Incl	4.15	1.875	36	27	396.2	1.5	29.88
7	7	N7	3	Incl	2.15	5	36	27	355.6	0.52	12.72
8	8	N8	9	Incl	4.15	5	36	42	373.4	0.69	17.16
9	9	N9	1	Incl	3.15	3.4375	31.5	34.5	377.6	0.97	23.31
10	10	N10	11	Incl	3.15	3.4375	31.5	34.5	381.2	1.05	21.78
11	11	N11	8	Incl	3.15	3.4375	31.5	34.5	376.5	0.95	20.45

Figure 13.12: *Experimental data of the first 11 laser welding experiments.*

We have now completed all steps of the problem formulation. Now the design must be executed and the response data entered into the worksheet. In the three right-most columns of the worksheet, numbered 9-11, we can see the response values found.

Fractional Factorial Designs

Introduction to fractional factorial designs

We have now completed the problem formulation part of the laser welding application, and carried out all experiments. This means that it is time for the data analysis. However, before we continue with that story, we will make a temporary stop and get more closely acquainted with the family of fractional factorial designs. For this purpose, we consider the 2^7 full factorial design in 128 runs. With 128 experiments it is possible to estimate 128 model

parameters, distributed as 1 constant term, 7 linear terms, 21 two-factor interactions, 35 three-factor interactions, 35 four-factor interactions, 21 five-factor interactions, 7 six-factor interactions, and 1 seven-factor interaction (Figure 13.13).

With the 2^7 full factorial design the following model parameters may be estimated:

constant	linear	2-fact. int.	3-fact. int.	4-fact. int.	5-fact. int.	6-fact. int.	7-fact. int.
1	7	21	35	35	21	7	1

Figure 13.13: *Estimable terms of the 2^7 full factorial design.*

Now, the fact that all these parameters can be estimated does not in any way guarantee that they are all of appreciable size and meaningful. Rather, there tends to be a certain hierarchy among model terms, making some terms more important than others. Looking at the absolute magnitude of model terms, we find that linear terms tend to be larger than two-factor interactions, which, in turn, tend to be larger than three-factor interactions, and so on. Consequently, it is often so that higher-order interactions tend to become negligible and can therefore be disregarded. Our long experience of applying DOE to chemical and technological problems suggests to us that three-factor interactions and interactions of higher order usually are negligible. This means that there tends to be a redundancy in a 2^k full factorial design, that is, an excess number of parameters which *can* be estimated but which lack relevance. This is the entry point for fractional factorial designs. They exploit this redundancy, by trying to reduce the number of necessary design runs.

A geometric representation of fractional factorial designs

The question which arises is *how is this decrease in the number of necessary experiments accomplished*? We begin by considering the geometry of the 2^3 full factorial design, as displayed in Figure 13.14. It represents the CakeMix application, except for the fact that all center-points are omitted from the drawing. With this full factorial design three factors are investigated in eight runs. Interestingly, the 2^3 full factorial design may be split into two balanced fractions, one shown in Figure 13.15 and one shown in Figure 13.16.
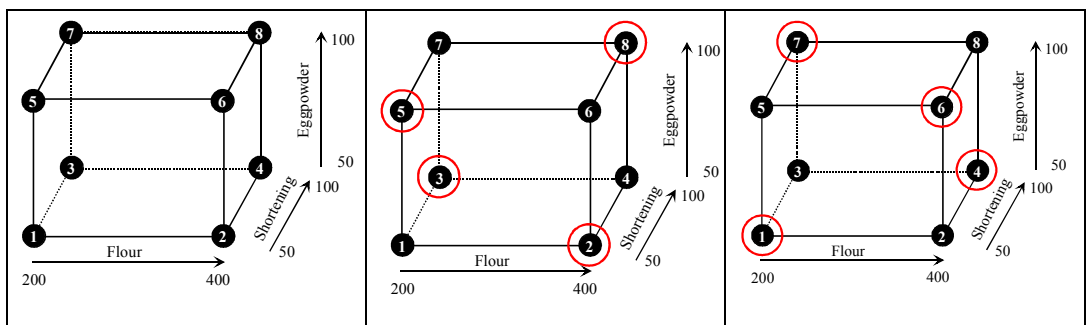
Figure 13.14: *(left) 2^3 full factorial design in eight runs.*
Figure 13.15: *(middle) First half-fraction of 2^3 design denoted 2^{3-1}.*
Figure 13.16: *(right) Second half-fraction of 2^3 design also denoted 2^{3-1}.*

The two half-fractions shown in Figures 13.15 and 13.16 are encoded by the so called 2^{3-1} fractional factorial design, and imply that three factors can be explored in four runs. From a design point of view these two fractions are equivalent, but in reality *one* may be preferred to the other because of practical experimental considerations. The fractional factorial design used in the laser welding application was created in an analogous fashion, only in this case

the 16 possible corners were divided into half-fractions each comprising eight experiments. These half-fractions are encoded by the 2^{4-1} fractional factorial design, and one of them was selected as the design to be performed.

In fact, by forming fractions like this, any two-level full factorial design may be converted into a fractional factorial design with a balanced distribution of experiments. However, with five or more factors, the parent two-level full factorial design may be reduced by more than one step. For instance, the 2^5 full factorial design may be reduced one step to become the 2^{5-1} fractional factorial design, or two steps to become the 2^{5-2} fractional factorial design. Using the former fractional factorial design, the 32 theoretically possible corner experiments are divided into two fractions of 16 experiments, and one such half-fraction is then selected as the working design. In the latter case, the 32 corners are divided into four fractions of eight experiments, and one of these quarter-fractions is then selected as the working design. Thus, in principle, it is possible to screen 5 factors in either 32, 16, or 8 experiments, plus some additional center-points. Which design to prefer will be discussed shortly.

Going from the 2^3 full factorial design to the 2^{4-1} fractional factorial design

We will now briefly describe the technique behind the fractionation of two-level full factorial designs. Consider the computational matrix of the 2^3 design shown in Figure 13.17. This design consists of eight corner experiments, making it possible to calculate a regression model of seven coefficients and one constant term. These seven regression coefficients represent 3 factor main effects, 3 two-factor interactions, and 1 three-factor interaction. Thus, we have eight experiments in the design and estimate eight model terms, which means that all possible effects are estimated.

Run #	constant	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	$x_1x_2x_3$
1	+	–	–	–	+	+	+	–
2	+	+	–	–	–	–	+	+
3	+	–	+	–	–	+	–	+
4	+	+	+	–	+	–	–	–
5	+	–	–	+	+	–	–	+
6	+	+	–	+	–	+	–	–
7	+	–	+	+	–	–	+	–
8	+	+	+	+	+	+	+	+

Figure 13.17: *Computational matrix of the 2^3 design.*

However, as was previously pointed out, three-factor interactions are usually negligible. This implies that in the 2^3 case the right-most column of the computational matrix encoding the $x_1x_2x_3$ three-factor interaction is not utilized fully. This is because it is set to represent a model term likely to be near-zero in numerical value. Hence, there is a slot "open" in the computational matrix for depicting a more prominent model term. As is shown in Figure 13.18, this free column may be assigned to another factor, say x_4. Thus, the column previously labeled $x_1x_2x_3$ now shows how to vary x_4 when doing the experiments. This way of introducing a fourth factor means that we have converted the 2^3 full factorial design to a 2^{4-1} fractional factorial design. Thus, we can explore four factors by means of eight

experiments, and not 16 experiments, as is the case with the 2^4 full factorial design. We will soon see what price we have to pay for this experiment reduction.

Run #	constant	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	$x_4 =$ $x_1x_2x_3$
1	+	−	−	−	+	+	+	−
2	+	+	−	−	−	−	+	+
3	+	−	+	−	−	+	−	+
4	+	+	+	−	+	−	−	−
5	+	−	−	+	+	−	−	+
6	+	+	−	+	−	+	−	−
7	+	−	+	+	−	−	+	−
8	+	+	+	+	+	+	+	+

Figure 13.18: How to introduce a fourth factor.

Confounding of effects

The price we have to pay when reducing the number of experiments is that our factor effects can no longer be computed completely free of one another. The effects are said to be *confounded*, that is, to a certain degree mixed up with each other. The extent to which the effects are confounded in the 2^{4-1} fractional factorial design, and, indirectly the laser welding application, is delineated in Figure 13.19. This figure contains a table of the confounding pattern, and there are two important observations to make. The first observation relates to the fact that the 16 possible effects are evenly allocated as two effects per column. The second observation is that the main effects are confounded with the three-factor interactions, and that the two-factor interactions are mutually confounded. This is a comparatively simple confounding situation, because all the effects of primary concern in screening, *the main effects*, are aliased with the negligible three-factor interactions. This means that when we calculate our regression model, and use the resulting regression coefficient plot for interpreting the model, we will be able to identify the most important factors with high certainty. This is addressed graphically in the ensuing section.

	$x_1x_2x_3x_4$	$x_2x_3x_4$	$x_1x_3x_4$	$x_1x_2x_4$	$= x_1x_2x_3$	x_3x_4	x_2x_4	x_2x_3
	constant	x_1	x_2	x_3	x_4	x_1x_2	x_1x_3	x_1x_4
1	+	−	−	−	−	+	+	+
2	+	+	−	−	+	−	−	+
3	+	−	+	−	+	−	+	−
4	+	+	+	−	−	+	−	−
5	+	−	−	+	+	+	−	−
6	+	+	−	+	−	−	+	−

Figure 13.19: Confounding pattern of the 2^{4-1} fractional factorial design.

A graphical interpretation of confoundings

We recall from the foregoing discussion that in the case of the 2^{4-1} fractional factorial design, the main effects are confounded with three-factor interactions and the two-factor interaction are confounded with each other. This means that when we try to interpret a regression model based on this design, we must bear in mind that the model terms are not completely resolved from one another. A schematic drawing of a regression model based on a 2^{4-1} fractional factorial design is shown in Figure 13.20. The confounding pattern of this model is listed underneath the coefficients.
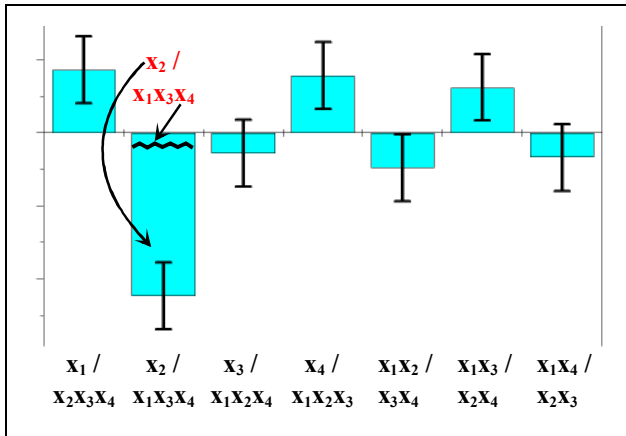
Figure 13.20: Hypothetical regression coefficients of a model based on a 2^{4-1} design.

Now, let us focus on the largest coefficient, the second one from the left. This regression coefficient represents the *sum* of the impact of the factor x_2 and the $x_1x_3x_4$ three-factor interaction. With the available experiments it is impossible to resolve these two terms. Fortunately, however, the three-factor interaction can be assumed to be of negligible relevance, and hence its contribution to the displayed regression coefficient is small. Thus, the regression coefficient that we see is a good approximation of the importance of x_2. A similar reasoning can of course be applied to the other main effects and three-factor interactions.

Concerning the confounding of the two-factor interactions, the situation is more complex and the interpretation harder. Sometimes it is possible to put forward an educated guess as to which two-factor interaction is likely to dominate in a confounded pair of two-factor interactions. Such reasoning is based on assessing which of the main effects are the largest ones. Factors with larger main effects normally contribute more strongly to a two-factor interaction than do factors with smaller main effects. Thus, regarding the set of regression coefficients given in Figure 13.20, it is reasonable to anticipate that in the confounded pair x_1x_3/x_2x_4, probably the x_2x_4 two-factor interaction contributes more strongly to the model. But the only way to be absolutely sure of this hypothesis is to do more experiments.

Resolution of fractional factorial designs

A confounding pattern in which main effects are aliased with three-factor interactions represents a workable trade-off between the number of factors screened and the number of experiments conducted. A measure of the complexity of a confounding pattern is given by the design's *resolution*. How to establish the resolution of a fractional factorial design is

clarified in Chapter 14. For the time being, we shall merely concentrate on how to interpret the resolution of a design.

Consider the summary displayed in Figure 13.21. This table provides an overview of the most commonly used fractional factorial designs and their respective resolutions. It is organized as follows: the vertical direction gives the number of corner experiments and the horizontal direction the number of treated factors. For each design, the resolution is given using roman numerals. Resolution III corresponds to a design in which main effects are confounded with two-factor interactions. This is undesirable in screening and hence resolution III designs should be used with great care. Resolution IV designs have main effects unconfounded with two-factor interactions, but the two-factor interactions are still confounded with each other. The 2^{4-1} fractional factorial design used in the laser welding example is of this kind. Resolution IV designs are recommended for screening, because they offer a suitable balance between the number of factors screened and the number of experiments needed. In designs of resolution V or higher, main effects are computed free of two-factor interactions, and the two-factor interactions themselves are unconfounded with each other. Resolution V and resolution V+ designs are almost as good as full factorial designs. Their main disadvantage is the large number of experiments required.

With this preliminary insight into design resolution, we are now ready to finish a previous discussion. Recall our discussion regarding whether five factors ought to be screened in 32, 16, or 8 experiments. The design overview table indicates that the resolution V 2^{5-1} design in 16 runs is the best choice, because there is no resolution IV design for five factors.

This concludes the introductory discussion concerning fractional factorial designs. The main advantage of fractional factorial designs is that many factors may be screened in drastically fewer runs. Their main drawback is the confounding of effects. However, to some extent, the confounding pattern may be regulated in a desired direction. With this discussion in mind we will, after a quiz break, return to the laser welding application, and the data analysis of this data-set.

Factors /Runs	3	4	5	6	7	8
4	2^{3-1} Res III +/- $X_3=X_1{*}X_2$	--	--	--	--	--
8	2^3	2^{4-1} Res IV +/- $X_4=X_1{*}X_2{*}X_3$	2^{5-2} Res III +/-$X_4=X_1{*}X_2$ +/-$X_5=X_1{*}X_3$	2^{6-3} Res III +/-$X_4=X_1{*}X_2$ +/-$X_5=X_1{*}X_3$ +/-$X_6=X_2{*}X_3$	2^{7-4} Res III +/-$X_4=X_1{*}X_2$ +/-$X_5=X_1{*}X_3$ +/-$X_6=X_2{*}X_3$ +/-$X_7=X_1{*}X_2{*}X_3$	--
16		2^4	2^{5-1} Res V +/- $X_5=X_1{*}X_2{*}X_3{*}X_4$	2^{6-2} Res IV +/-$X_5=X_1{*}X_2{*}X_3$ +/-$X_6=X_2{*}X_3{*}X_4$	2^{7-3} Res IV +/-$X_5=X_1{*}X_2{*}X_3$ +/-$X_6=X_2{*}X_3{*}X_4$ +/-$X_7=X_1{*}X_3{*}X_4$	2^{8-4} Res IV +/-$X_5=X_2{*}X_3{*}X_4$ +/-$X_6=X_1{*}X_3{*}X_4$ +/-$X_7=X_1{*}X_2{*}X_3$ +/-$X_8=X_1{*}X_2{*}X_4$
32			2^5	2^{6-1} D-opt Res VI +/-$X_6=X_1{*}X_2{*}X_3$ $*X_4{*}X_5$	2^{7-2} D-opt Res IV +/-$X_6=X_1{*}X_2{*}X_3{*}X_4$ +/-$X_7=X_1{*}X_2{*}X_4{*}X_5$	2^{8-3} Res IV +/-$X_6=X_1{*}X_2{*}X_3$ +/-$X_7=X_1{*}X_2{*}X_4$ +/-$X_8=X_2{*}X_3{*}X_4{*}X_5$

Figure 13.21: Overview of the resolution of some common fractional factorial designs.

Quiz

Please answer the following questions:

How many factors is it rational to screen at the same time?

Which important questions regarding the factors are usually asked in screening?

What is an Ishikawa diagram? In which respect is it helpful?

Which types of regression model are normally utilized in screening?

What is a fractional factorial design?

How are fractional factorial designs constructed (describe in simple terms)?

How many corner experiments are encoded by a 2^{4-1} fractional factorial design? By a 2^{5-1} design? By a 2^{5-2} design?

What is meant by confounding of effects?

What does the resolution of a design signify?

Which resolution is recommended for screening?

Laser welding application I

Evaluation of raw data

We start the evaluation of the raw data by inspecting the replicate error of each response. The three relevant plots of replications are displayed in Figures 13.22 – 13.24. As seen, the replicate errors are small for all three responses, which means that we have good data to work with.
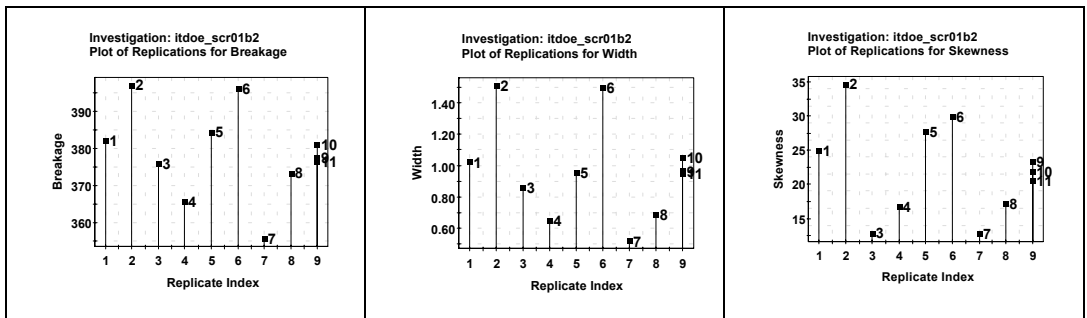
Figure 13.22: (left) Replicate plot of Breakage.
Figure 13.23: (middle) Replicate plot of Width.
Figure 13.24: (right) Replicate plot of Skewness.

Further, in the evaluation of the raw data, it is mandatory to check the data distribution of the responses, to reveal any need for response transformations. We may check this by making a histogram of each response. Such histograms are rendered in Figures 13.25 – 13.27, and they inform us that it is pertinent to work in the untransformed metric of each response.
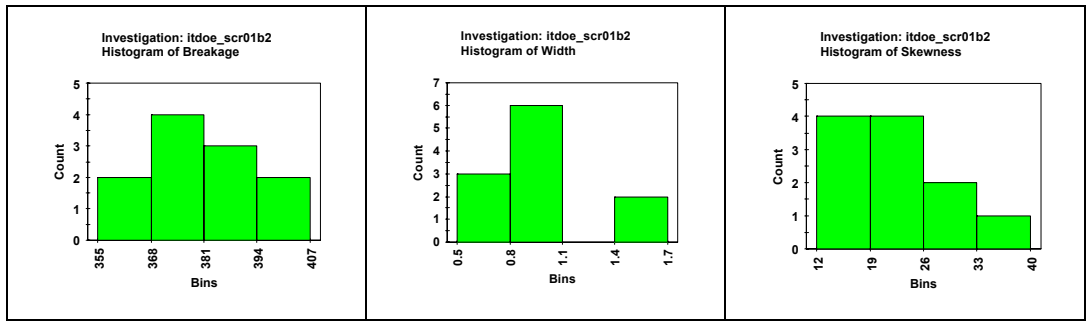
Figure 13.25: *(left) Histogram of Breakage.*
Figure 13.26: *(middle) Histogram of Width.*
Figure 13.27: *(right) Histogram of Skewness.*

Other aspects of the raw data that might be worthwhile to consider are the condition number of the fractional factorial design, and the inter-relatedness among the three responses. We notice from Figure 13.28 that the condition number is approximately 1.2, which lies within the interval of 1 – 3 in which standard screening designs typically are found. The correlation matrix listed in Figure 13.29 tells us that the three responses are strongly correlated, because their correlation coefficients range from 0.88 to 0.96. Hence, fitting three MLR models with identical composition of model terms appears reasonable.
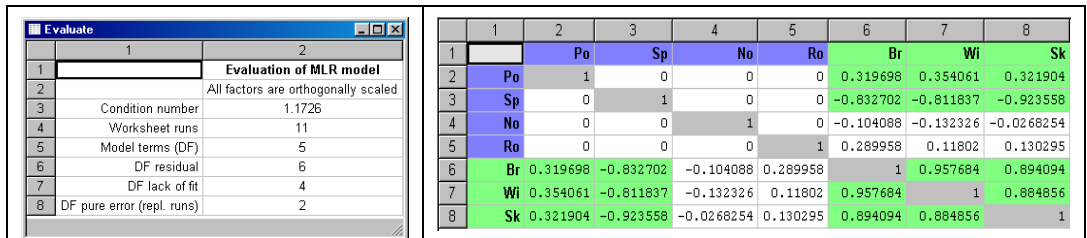
Figure 13.28: *(left) Condition number evaluation of laser welding design.*
Figure 13.29: *(right) Plot of correlation matrix of laser welding design. Note the high inter-relatedness among the three responses.*

Regression analysis

We will now fit a linear model with five terms, the constant and four linear terms, to each response. The overall results of the model fitting are summarized in Figure 13.30. Evidently, the predictive power, as evidenced by Q^2, is acceptable only for Skewness. The prediction ability is poor for Breakage and Width. In trying to understand why this might be the case, it is appropriate to continue with the second diagnostic tool, the analysis of variance. ANOVA tables of the three responses are given in Figures 13.31 – 13.33. Remembering that the upper p-value should be smaller than 0.05 and the lower p-value larger than 0.05, we see that all three models pass this test.
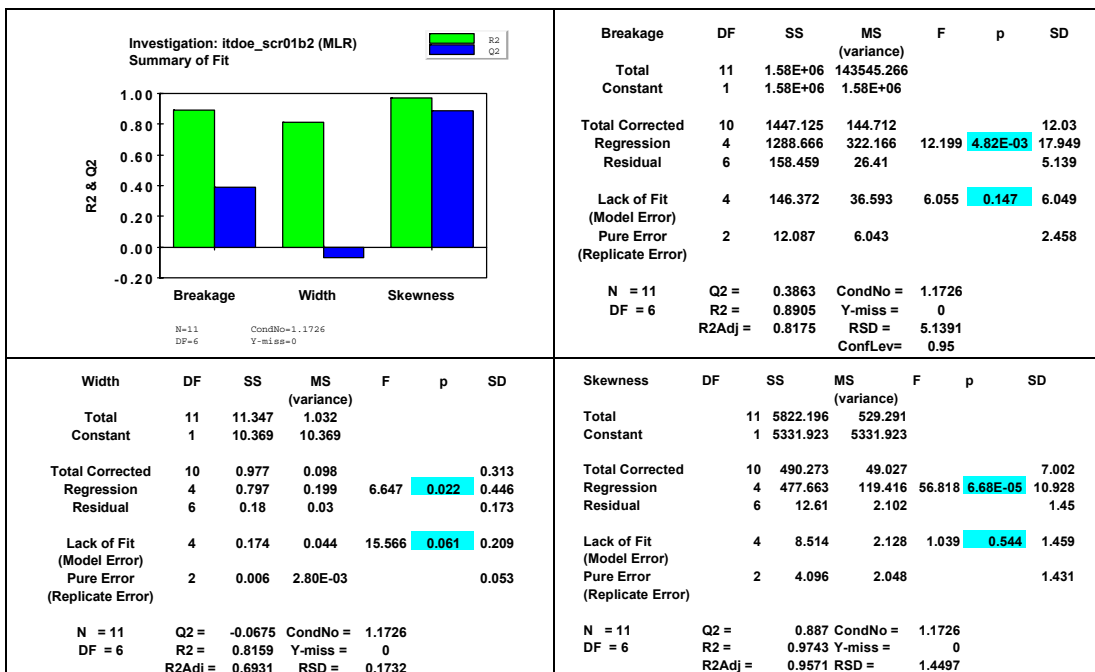
Figure 13.30: *(upper left) Summary of fit plot of laser welding model.*
Figure 13.31: *(upper right) ANOVA of Breakage.*
Figure 13.32: *(lower left) ANOVA of Width.*
Figure 13.33: *(lower right) ANOVA of Skewness.*

The outcome of the third diagnostic tool, the normal probability plot of residuals, is shown in Figures 13.34 – 13.36. These plots indicate that the residuals for experiments 3 and 6 are unusually large, but we should not delete any experiment.
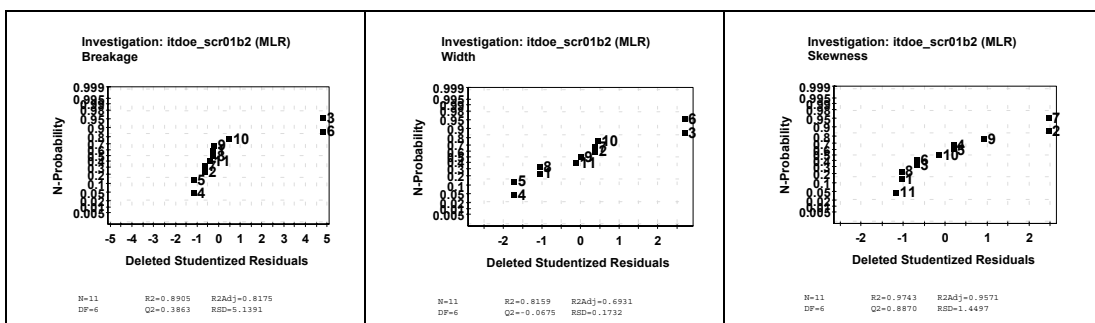
Figure 13.34: *(left) Normal probability plot of Breakage.*
Figure 13.35: *(middle) Normal probability plot of Width.*
Figure 13.36: *(right) Normal probability plot of Skewness.*

Since we have not yet obtained any tangible information as to why Q^2 is low for Breakage and Width, it is appropriate to consult plots of regression coefficients. Such regression coefficient plots are shown in Figures 13.37 – 13.40. The last of these is an overview of all regression coefficients, aimed at facilitating the simultaneous interpretation of several models. Apparently, two factors dominate, that is, Power and Speed of laser. The third

factor, NozzleGas, does not influence any response. The fourth factor, RootGas, exerts a minor influence with regards to Breakage. Based on this information, it appears reasonable to try to modify the original model by adding the two-factor interaction Po*Sp. However, in so doing, it must be remembered that this model term is confounded with another two-factor interaction, the one denoted No*Ro.
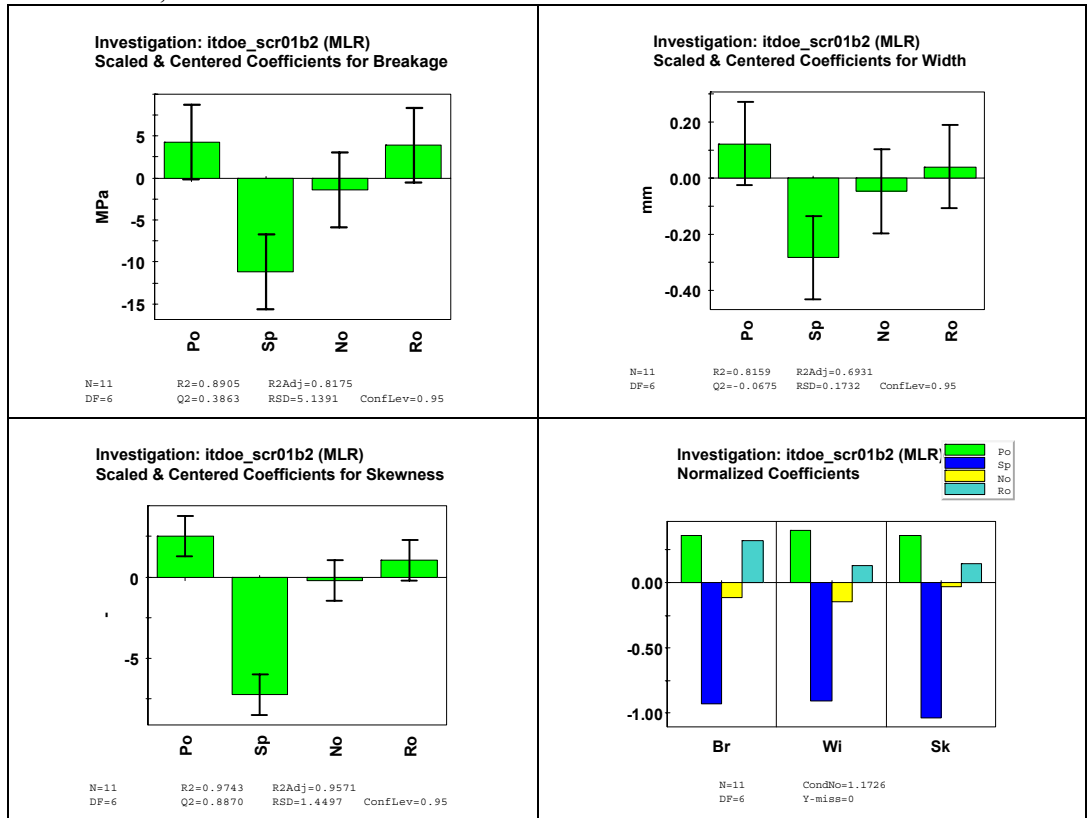
Figure 13.37: *(upper left) Regression coefficients for Breakage.*
Figure 13.38: *(upper right) Regression coefficients for Width.*
Figure 13.39: *(lower left) Regression coefficients for Skewness.*
Figure 13.40: *(lower right) Overview of regression coefficients.*

Model refinement

We will now carry out a refinement of the linear model. For reference purposes we re-display in Figures 13.41 and 13.42 the summary of fit plot and the coefficient overview plot of the original model. The progress of the model pruning may easily be overviewed using similar pairs of graphs.
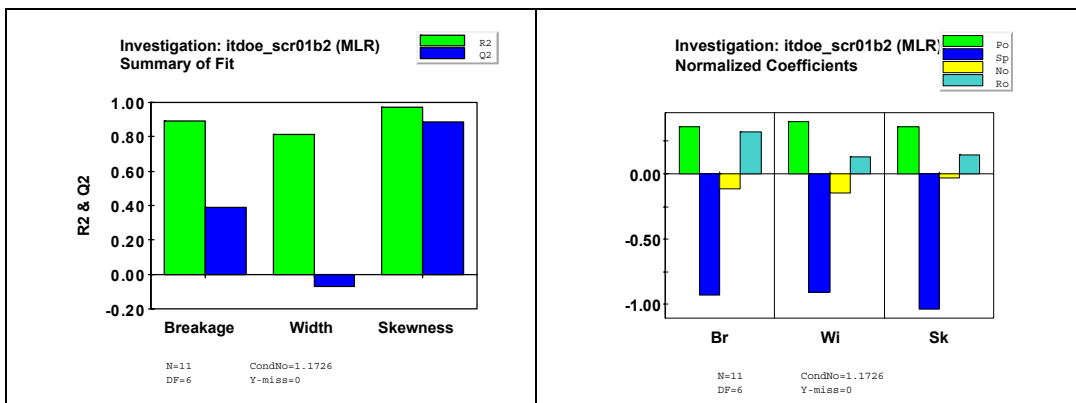
Figure 13.41: *(left) Summary of fit of original model.*
Figure 13.42: *(right) Coefficient overview of original model.*

In the first step, the Po*Sp two-factor interaction was incorporated, and the model refitted. As is evident from Figures 13.43 and 13.44, the inclusion of the interaction term is mainly beneficial for the second response, but its presence reduces the predictive power regarding the first response. We also observe that the linear term of NozzleGas is not influential in the model. Hence, it is relevant to remove this term and evaluate the consequences.
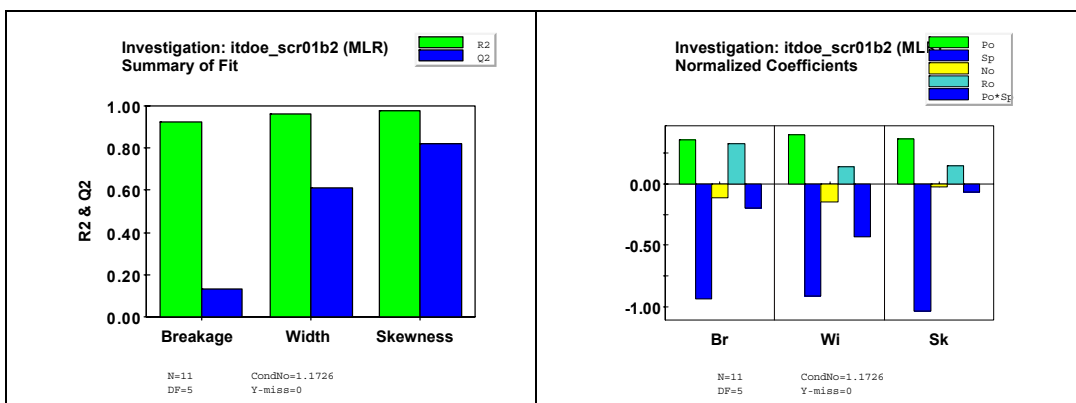
Figure 13.43: *(left) Summary of fit after inclusion of Po*Sp*
Figure 13.44: *(right) Coefficient overview after inclusion of Po*Sp.*

Figures 13.45 and 13.46 show that all three regression models are enhanced as a result of the exclusion of NozzleGas. The model summary shown in Figure 13.45 represents the "best" modelling situation that we have obtained. In general, a Q^2 above 0.5 is good and above 0.9 is excellent. Thus, Figure 13.45 demonstrates that we are doing a good job with regards to Skewness, but are doing less well with respect to Breakage. The gap between R^2 and Q^2 concerning Breakage is unsatisfactory. The modelling of Width is acceptable for screening purposes.
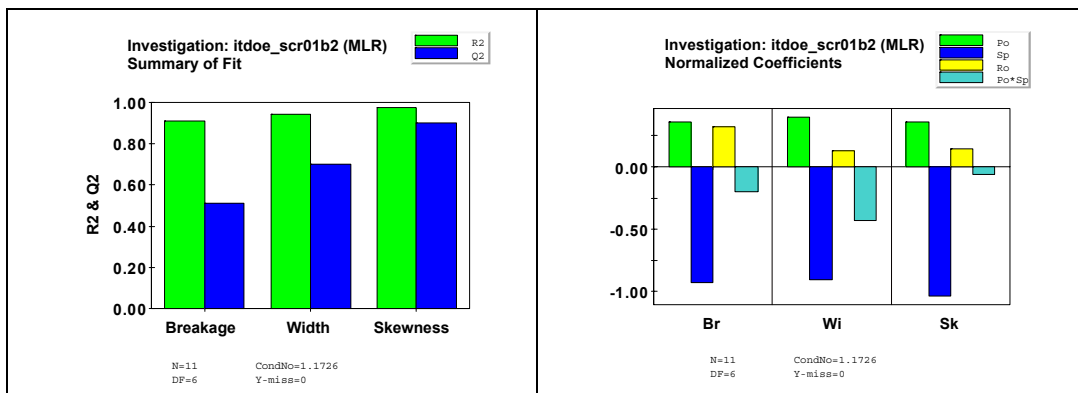
Figure 13.45: *(left) Summary of fit after exclusion of NozzleGas.*
Figure 13.46: *(right) Coefficient overview after exclusion of NozzleGas.*

We may now proceed by reviewing the other diagnostic tools, the results of which are exposed in Figures 13.47 – 13.52. The ANOVA examination reveals that all models are acceptable from that perspective, that is, they show no lack of fit. The N-plots of residuals indicate that one plausible reason for the low Q^2's of Breakage and Width might be the deviating behavior of experiments 3 and 7. It may well be the case that some additional two-factor interaction is needed to better account for this deviation.

Breakage	DF	SS	MS (variance)	F	p	SD
Total	11	1.58E+06	143545.27			
Constant	1	1.58E+06	1.58E+06			
Total Corrected	10	1447.125	144.712			12.03
Regression	4	1319.065	329.766	15.451	2.59E-03	18.159
Residual	6	128.06	21.343			4.62
Lack of Fit (Model Error)	4	115.973	28.993	4.798	0.18	5.385
Pure Error (Replicate Error)	2	12.087	6.043			2.458

N = 11	Q2 =	0.5118	CondNo =	1.1726
DF = 6	R2 =	0.9115	Y-miss =	0
	R2Adj =	0.8525	RSD =	4.6199
			ConfLev=	0.95

Investigation: itdoe_scr01b2 (MLR)
Breakage

N=11   R2=0.9115   R2Adj=0.8525
DF=6   Q2=0.5118   RSD=4.6199

Width	DF	SS	MS (variance)	F	p	SD
Total	11	11.347	1.032			
Constant	1	10.369	10.369			
Total Corrected	10	0.977	0.098			0.313
Regression	4	0.923	0.231	25.677	6.45E-04	0.48
Residual	6	0.054	0.009			0.095
Lack of Fit (Model Error)	4	0.048	0.012	4.316	0.197	0.11
Pure Error (Replicate Error)	2	0.006	2.80E-03			0.053

N = 11	Q2 =	0.7028	CondNo =	1.1726
DF = 6	R2 =	0.9448	Y-miss =	0
	R2Adj =	0.908	RSD =	0.0948

Investigation: itdoe_scr01b2 (MLR)
Width

N=11   R2=0.9448   R2Adj=0.9080
DF=6   Q2=0.7028   RSD=0.0948

Skewness	DF	SS	MS (variance)	F	p	SD
Total	11	5.82E+03	529.291			
Constant	1	5.33E+03	5.33E+03			
Total Corrected	10	490.273	49.027			7.002
Regression	4	478.93	119.733	63.332	4.87E-05	10.942
Residual	6	11.343	1.891			1.375
Lack of Fit (Model Error)	4	7.247	1.812	0.885	0.592	1.346
Pure Error (Replicate Error)	2	4.096	2.048			1.431

N = 11	Q2 =	0.9024	CondNo =	1.1726
DF = 6	R2 =	0.9769	Y-miss =	0
	R2Adj =	0.9614	RSD =	1.375

Figure 13.47: (upper left) ANOVA of Breakage – after model refinement.
Figure 13.48: (upper right) N-plot of residuals of Breakage – after model refinement.
Figure 13.49: (middle left) ANOVA of Width – after model refinement.
Figure 13.50: (middle right) N-plot of residuals of Width – after model refinement.
Figure 13.51: (lower left) ANOVA of Skewness – after model refinement.
Figure 13.52: (lower right) N-plot of residuals of Skewness – after model refinement.

Because of the comparatively poor modelling of the first response, and the suspected need for more two-factor interactions, the investigator decided to carry out more experiments in order to upgrade the underlying design and, thereby, also the regression model. To perform this upgrade the investigator used a technique called fold-over. Fold-over will be explained in a moment.

Use of model

Although it has already been decided to do more experiments by using the fold-over technique, it is useful to interpret and use the existing model. This we do primarily to determine whether the desired response profile is likely to be realized within the investigated experimental region. Since the "working" regression model contains the Po*Sp two-factor interaction, it is logical to construct response contour plots focusing on this term. Such a triple-contour plot is shown in Figure 13.53.
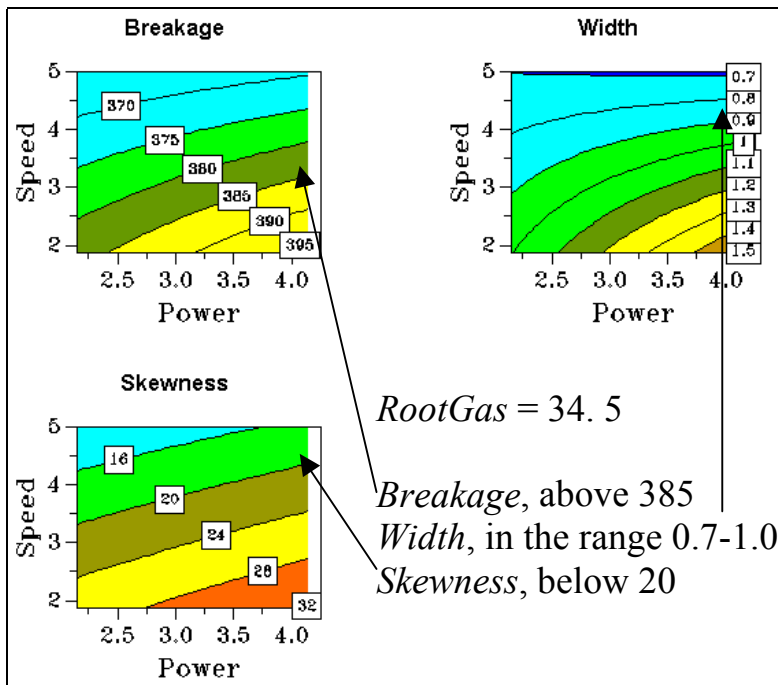
Figure 13.53: *Triple-contour plot of Breakage, Width, and Skewness.*

We can see how the three responses change as a result of changing the factors Speed and Power, while keeping RootGas fixed at its center level. Observe that NozzleGas is not included in the model, and hence does not affect the shape of these plots. The arrows indicate where the stated goals are fulfilled with regard to each individual response. It is evident that no unique point exists in these contours where all desirabilities are fulfilled. However, staying somewhere along the right-hand edge seems appropriate. Perhaps after the inclusion of the 11 additional experiments it will be possible to pinpoint an optimum. This remains to be seen. Later on when all 22 runs are analyzed together, we will use an automatic search functionality, an optimizer, to locate more precisely a possible optimum.

This concludes the first stage in the analysis of the laser welding data. We will now introduce some more theory relating to what to do after screening, as this matches our current situation. Then we will revert to the data analysis.

What to do after screening

Introduction

Exactly what to do after a screening investigation depends on a number of things. By necessity, this decision will be influenced by the quality of the regression model obtained, whether it is possible and/or necessary to modify the factor ranges, and whether some of the experiments already conducted are close to fulfilling the goals stated in the problem formulation.

In the broadest sense, the measures to be taken after screening can be divided in two major approaches. One approach is based on changing the factor ranges and thus moving the

experimental region. The other approach is based on *not* altering the factor settings. We will commence by discussing the latter approach, and in particular what options are available for further exploration of an already mapped experimental region. Our treatment of how to act when moving an experimental region will be given later.

Experiments outside investigated region impossible or undesirable

There are three situations in which performing experiments outside the explored experimental region is either impossible or undesirable. The first situation corresponds to the ideal outcome of a screening. This occurs when one of the performed experiments fulfills the goal of the problem formulation. In this case, experimenting outside the explored region is undesirable, and all we need to do is a couple of verifying experiments.

The second situation, which is more problematic, but often manageable, occurs when none of the performed experiments meets our goals, but it is not possible to experiment outside the specified factor ranges. In this case, we must rely on our screening model for predicting the best possible experimental point. Of course, this can only be done with a reliable model, as predictions made with a poor model may be terribly misleading. Subsequently, we must carry out an experiment at this predicted best point to verify that the response profile is reasonably close to the stated goals.

The third situation occurs when it is suspected that the model is too simple for the problem, or that a poor model has been obtained which needs to be improved. The laser welding application is a good example of this last situation. Here, we need a better model for Breakage and we suspect that an interaction model would do a better job than the fitted, slightly updated, linear model. This third situation is best addressed by extending the parent design with some complementary experimental runs. We will address this in the next section.

Adding complementary runs

There are two primary motives for adding complementary runs to a design, namely complementing for unconfounding and complementing for curvature. Fractional factorial designs of resolution III and IV have confoundings among main effects and two-factor interactions. This means that these effects can not be computed clear of one another. It is possible to resolve confounded effects, that is, to unconfound, by adding extra experiments.

When *complementing for unconfounding* purposes, one may either (i) upgrade the whole design and the accompanying model, or (ii) selectively update the design with a minimum number of runs for unconfounding some specific effects that are of interest. Upgrading the entire design is usually accomplished with a technique called *fold-over*. However, this technique is unselective. It enables all terms of a certain complexity to be better resolved, and is therefore quite costly in terms of additional experiments. If a clear picture exists that, for example, two two-factor interactions need to be estimated free of each other, it is sufficient to selectively update the design with two additional experiments, rather than a bunch of new experiments. Such a selective design supplement is usually accomplished with a D-optimal design, but this an advanced topic.

The second main reason for adding complementary experiments occurs when there is a need to model curvature, that is, non-linear relationships among factors and responses. Fractional and full factorial designs may be complemented by (i) upgrading the entire design to a

composite design, or by (ii) selective updating, as outlined above. In both instances, full or partial quadratic models result, which are capable of handling curved relationships. Composite designs are discussed in detail in connection with General Example 2 in Chapter 15.

Fold-over

The fold-over design is a complementary design that is usually added to achieve unconfounding in resolution III and resolution IV designs. It is a collection of experiments selected from the other existing, but unused, fractions. This is exemplified in Figure 13.54. The upper eight experiments originate from the 2^{5-2} fractional factorial design. The fold-over complement of this resolution III design is given by the lower eight experiments. When comparing the sign pattern of the upper quarter-fraction with that of the lower quarter-fraction, we can see that they have identical sequences except for a sign inversion. The fold-over complement was selected from all possible quarter-fractions, by switching the sign of the two columns used to introduce the fourth and the fifth factors.

x1	x2	x3	x4 = x1x2	x5 = x1x3
-	-	-	+	+
+	-	-	-	-
-	+	-	-	+
+	+	-	+	-
-	-	+	+	-
+	-	+	-	+
-	+	+	-	-
+	+	+	+	+
+	+	+	-	-
-	+	+	+	+
+	-	+	+	-
-	-	+	-	+
+	+	-	-	+
-	+	-	+	-
+	-	-	+	+
-	-	-	-	-
			x4 = -x1x2	x5 = -x1x3

Figure 13.54: The 2^{5-2} design and its fold-over.

When the initial design is of resolution III, as is the case with the 2^{5-2} fractional factorial design, the initial design together with its fold-over will be of resolution IV. This means that all main effects will be resolved from the two-factor interactions. Note, however, that this design is not as efficient as the 2^{5-1} fractional factorial design with resolution V. Moreover, when the original design is of resolution IV, the original design and its fold-over only

sometimes become a resolution V design. MODDE automatically takes care of folding over a fractional factorial design, when this command is launched.

Creating the fold-over of the laser welding screening design

Because the starting design in the laser welding application was a resolution IV 2^{4-1} fractional factorial design, this design combined with its fold-over will be the complete 2^4 factorial design. The upgraded design is shown in Figure 13.55 in terms of the new worksheet. We can see that the worksheet has been appended with 11 new experiments, found in rows 12 through 22. Eight new corner experiments have been conducted, thus completing the factorial part of the design, as well as three additional center-points. Thus, we now have six replicated center-points.

	Exp No	Exp Name	Run Order	Incl/Excl	Power	Speed	NozzleGas	RootGas	$Block	Breakage	Width	Skewness
1	1	N1	5	Incl	2.15	1.875	27	27	−1	382.2	1.02	24.96
2	2	N2	4	Incl	4.15	1.875	27	42	−1	397.2	1.51	34.68
3	3	N3	2	Incl	2.15	5	27	42	−1	375.8	0.86	12.84
4	4	N4	7	Incl	4.15	5	27	27	−1	365.6	0.65	16.68
5	5	N5	10	Incl	2.15	1.875	36	42	−1	384.4	0.96	27.72
6	6	N6	6	Incl	4.15	1.875	36	27	−1	396.2	1.5	29.88
7	7	N7	3	Incl	2.15	5	36	27	−1	355.6	0.52	12.72
8	8	N8	9	Incl	4.15	5	36	42	−1	373.4	0.69	17.16
9	9	N9	1	Incl	3.15	3.4375	31.5	34.5	−1	377.6	0.97	23.31
10	10	N10	11	Incl	3.15	3.4375	31.5	34.5	−1	381.2	1.05	21.78
11	11	N11	8	Incl	3.15	3.4375	31.5	34.5	−1	376.5	0.95	20.45
12	12	New12	12	Incl	2.15	1.875	27	42	1	383.2	1.01	22.2
13	13	New13	18	Incl	4.15	1.875	27	27	1	397.2	1.41	65.4
14	14	New14	19	Incl	2.15	5	27	27	1	362.2	0.67	12.12
15	15	New15	22	Incl	4.15	5	27	42	1	369.6	0.66	21.48
16	16	New16	16	Incl	2.15	1.875	36	27	1	383.8	0.94	23.88
17	17	New17	15	Incl	4.15	1.875	36	42	1	402.4	1.53	32.04
18	18	New18	17	Incl	2.15	5	36	42	1	365.8	0.61	14.76
19	19	New19	20	Incl	4.15	5	36	27	1	368	0.73	14.4
20	20	New20	14	Incl	3.15	3.4375	31.5	34.5	1	374.6	1.06	24.3
21	21	New21	21	Incl	3.15	3.4375	31.5	34.5	1	382.9	0.89	22.1
22	22	New22	13	Incl	3.15	3.4375	31.5	34.5	1	378	0.93	26.8

Figure 13.55: *The new laser welding worksheet after fold-over.*

In addition, the software has added a block factor. This is a precautionary measure that is useful for probing whether significant changes over time have occurred in the response data. If small or near-zero, this block factor may be removed from the model, and hence the resulting design can be regarded as the 2^4 full factorial design. If the block factor is significant, the parent design combined with its fold-over corresponds to a fractional factorial design of resolution V+.

Laser welding application II

Evaluation of raw data II

We see from the three replicate plots in Figures 13.56 – 13.58, that the replicate error is small for every response. Because of the presence of the block factor, the two triplets of replicated center-points are recognized as different and therefore rendered on separate bars. We will evaluate in the data analysis whether the block term is essential or not. If it is small

and insignificant, we have, in principle, no drift over time and in that case we have six
identically replicated center-points in the design.
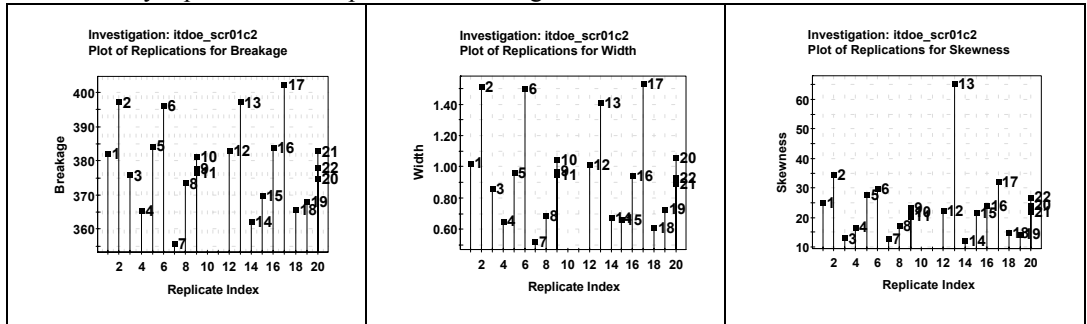
Figure 13.56: *(left) Replicate plot of Breakage.*
Figure 13.57: *(middle) Replicate plot of Width.*
Figure 13.58: *(right) Replicate plot of Skewness.*

Another noteworthy observation is the extreme value of trial #13 in the Skewness response.
It indicates that this response is not normally distributed. Indeed, the three histograms of
Figures 13.59 – 13.61 indicate that this is the case. Hence, it was decided to transform
Skewness. The histogram in Figure 13.62 shows the situation after log-transforming
Skewness. Finally, the condition number evaluation shows that the condition number of the
design is nearly 1.2. No plot is shown to illustrate this.
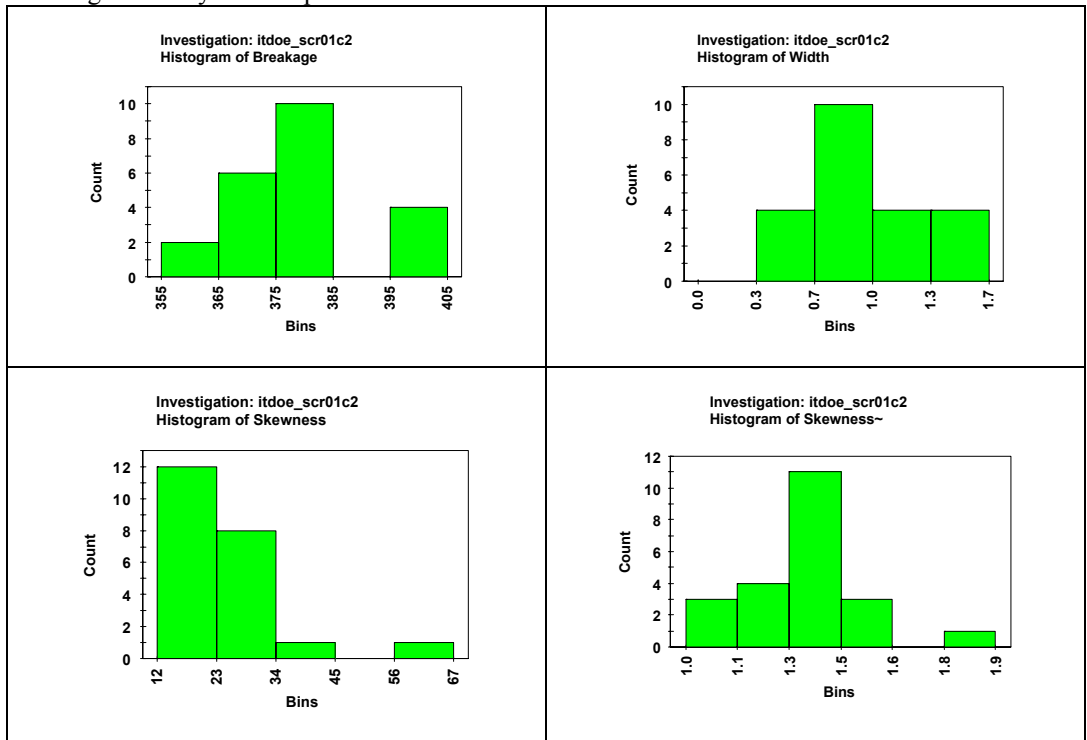
Figure 13.59: *(upper left) Histogram of Breakage.*
Figure 13.60: *(upper right) Histogram of Width.*
Figure 13.61: *(lower left) Histogram of Skewness prior to transformation.*
Figure 13.62: *(lower right) Histogram of Skewness after transformation.*

Regression analysis II

The result of fitting an interaction model with 15 terms to the three responses is shown in Figure 13.63. Evidently, we have obtained good models for Breakage and Width, but not for Skewness. The ANOVA's of these models are all acceptable, as no lack of fit is detected. We provide no plots related to the ANOVA's.
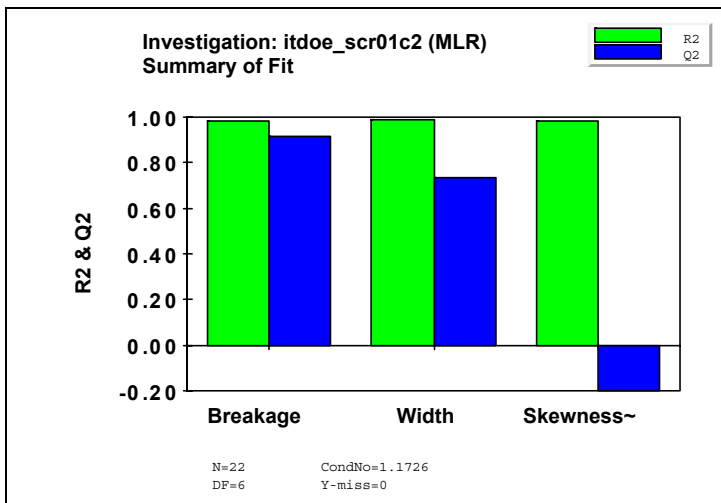
Figure 13.63: *Summary of fit of the interaction model.*

However, the N-plots of residuals, in Figures 13.64 – 13.66, display strange residual patterns. Although all points are located well within ± 3 standard deviations, the layered appearance of the points is conspicuous and should be more closely inspected.
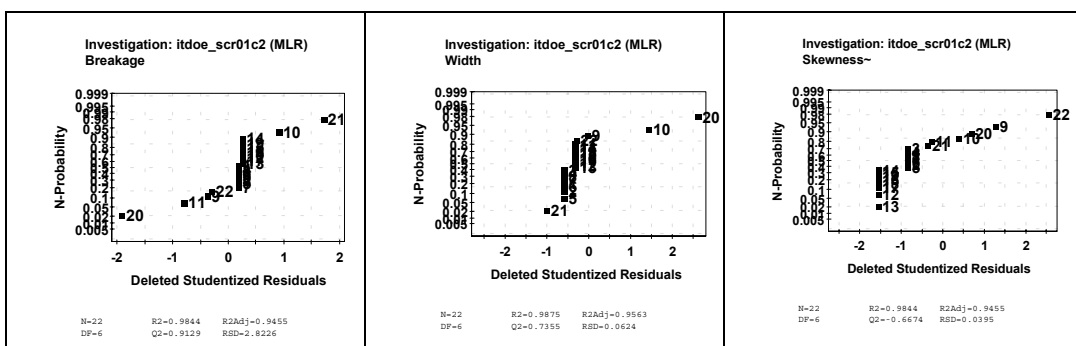
Figure 13.64: *(left) N-plot of residuals of Breakage (interaction model).*
Figure 13.65: *(middle) N-plot of residuals of Width (interaction model).*
Figure 13.66: *(right) N-plot of residuals of Skewness (interaction model).*

Next, we proceed to the plots of regression coefficients provided in Figures 13.67 – 13.69. After some scrutiny of these plots, it is obvious that the block factor is insignificant and should be deleted from the modelling. This indicates that there is no systematic drift between the two points in time at which the parent design and its fold-over were conducted. But it is not unusual that, for instance, variations between batches of raw materials, changes in ambient temperature and moisture, or other uncontrollable factors, may induce significant

shifts in the response data. Furthermore, from an overall model interpretation viewpoint, all two-factor interactions except Po*Sp and Po*No are small and may be removed. Thus, it is sensible to delete nine model terms: the block factor and eight two-factor interactions. This is discussed further in the next section.

Figure 13.67: (left) Regression coefficients of Breakage (interaction model).
Figure 13.68: (middle) Regression coefficients of Width (interaction model).
Figure 13.69: (right) Regression coefficients of Skewness (interaction model).
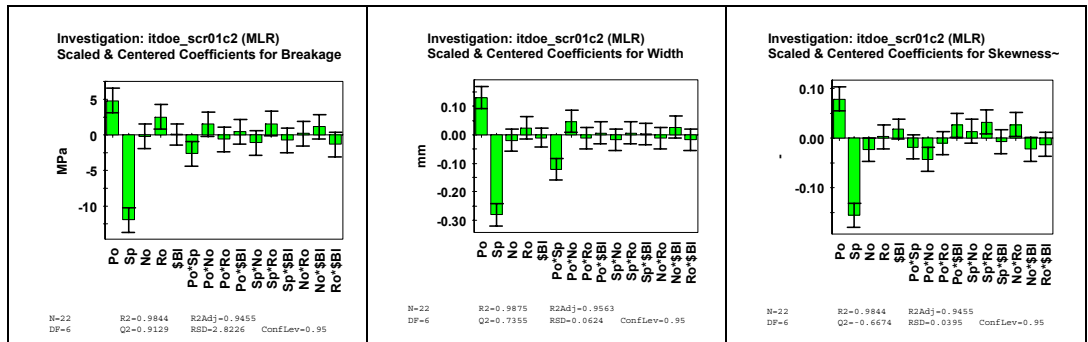
Model refinement II

Upon removal of the 9 model terms and refitting the three regression models, the summary statistics given in Figure 13.70 were obtained. We now have Q^2's of 0.86, 0.94, and 0.68 for Breakage, Width, and Skewness, respectively. These values range from good to excellent. Hence, the pruning of the model has been beneficial. The ANOVA tables in Figures 13.71 – 13.73 reveal the adequacies of the three models. No lack of fit is detected.

Figure 13.70: (upper left) Summary of fit of the refined interaction model.
Figure 13.71: (upper right) ANOVA of the refined interaction model of Breakage.
Figure 13.72: (lower left) ANOVA of the refined interaction model of Width.
Figure 13.73: (lower right) ANOVA of the refined interaction model of Skewness.
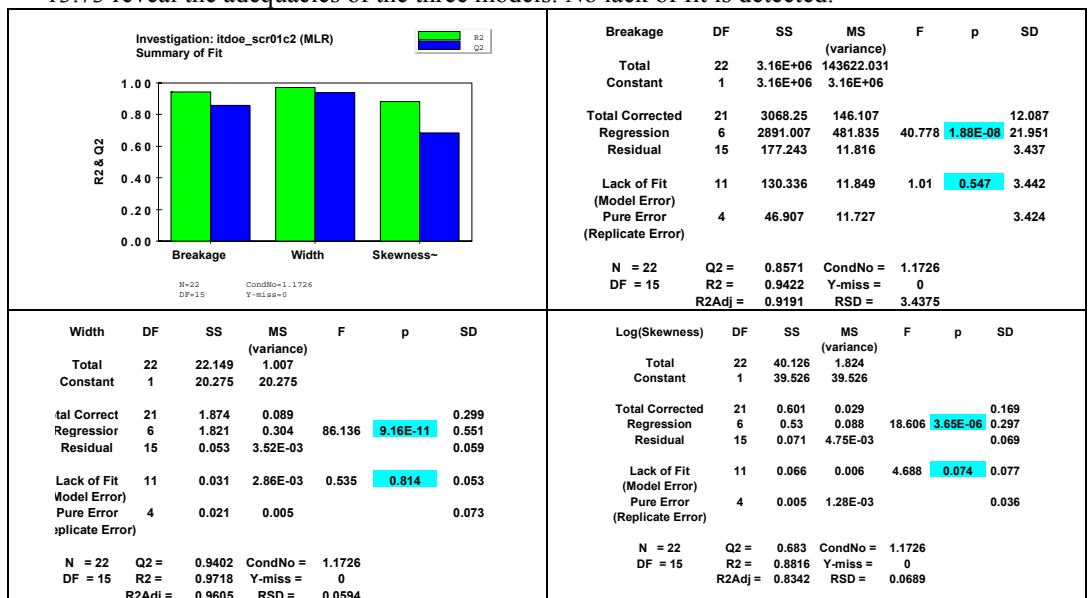
Moreover, the N-plots of residuals in Figures 13.74 – 13.76 are now much smoother. However, we have an indication that experiment #13 is deviating with respect to Skewness. This is the extreme value of this response, but it is a weak outlier. We define it as a weak outlier because Q^2 is relatively high at 0.68, and no lack of fit is discernible. However, if this experiment is removed and the Skewness model again refitted, the Q^2 for Skewness raises from 0.68 to 0.84. This constitutes some evidence that the presence of #13 slightly degrades Q^2 for Skewness. However, because #13 is well accounted for in the modelling of Breakage and Width, we decided to keep it in our model for Skewness as well.
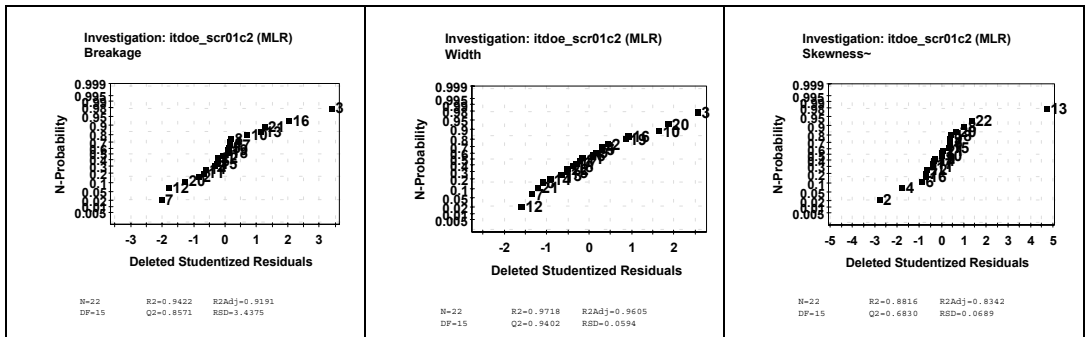
Figure 13.74: *(left) N-plot of the refined interaction model of Breakage.*
Figure 13.75: *(middle) N-plot of the refined interaction model of Width.*
Figure 13.76: *(right) N-plot of the refined interaction model of Skewness.*

The regression coefficients of the three models are given in Figures 13.77 – 13.79. Because the coefficient profiles are rather similar for all three responses, we may infer that the responses are correlated. A quick glimpse into the correlation matrix (no plot shown) corroborates this, as the correlation coefficients between the responses vary between 0.82 and 0.95. The two most important factors are the Power and Speed of the laser. NozzleGas is only of limited relevance, but it is kept in the final model since it participates in the Po*No two-factor interaction. The last factor, RootGas, is most meaningful for Breakage. Given that all three responses need to be well modelled at the same time, these three models represent the best compromise we can obtain. We will now use these models in trying to predict an optimal factor setting.
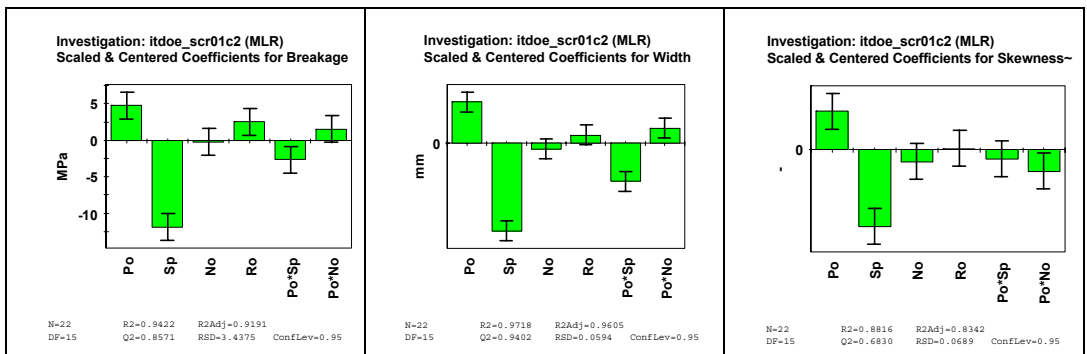
Figure 13.77: *(left) Regression coefficients of the refined interaction model of Breakage.*
Figure 13.78: *(middle) Regression coefficients of the refined interaction model of Width.*
Figure 13.79: *(right) Regression coefficients of the refined interaction model of Skewness.*

Use of model II

Having acquired technically sound regression models of Breakage, Width, and Skewness, the next important step in the DOE process corresponds to identifying a factor combination at which all goals are likely to be achieved. This process is greatly facilitated by response contour plots, or response surface plots. An interesting triplet of response contour plots pertaining to the plane defined by Speed and Power, and keeping NozzleGas and RootGas fixed at their center-levels, is displayed in Figure 13.80. Alongside these plots the goals of the three responses are listed.
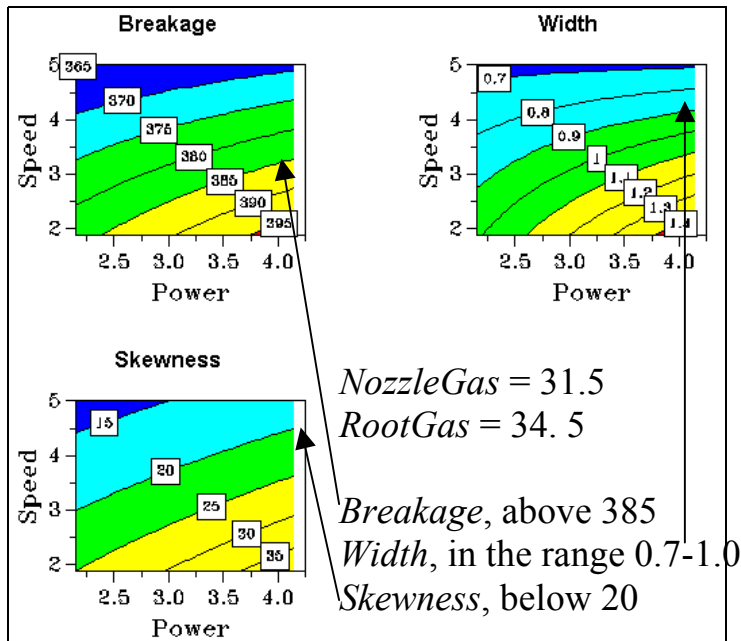
Figure 13.80: Triple-contour plot of Breakage, Width, and Skewness, refined interaction model.

The three arrows suggest that we should stick to the right-hand edge, but it is evident that there is no unique point at which all the goals are met simultaneously. In this situation, it is necessary to browse through many other such sets of contour triplets to identify an optimal point. However, this is both laborious and time-consuming, and there is really no guarantee that an optimal point will be found inside the investigated domain. Hence, it might be necessary to explore new territories, outside the mapped domain. Fortunately, two helpful techniques are available for this, which we intend to describe.

Experiments outside investigated region possible and/or desirable

We are now in the situation where we want to use the derived regression models. And we want to use them to predict where to do our next experiment, which hopefully will correspond to, or be in the vicinity of, the optimal point. This means that we are just about to confront one of the minor experimental objectives, the one entitled *finding the optimal region*. Finding the optimal region is an objective which is often used to bridge the gap between screening and optimization. As seen in Figure 13.81, we want to understand how to alter the factor settings so that we enter into an area which does include the optimal point.
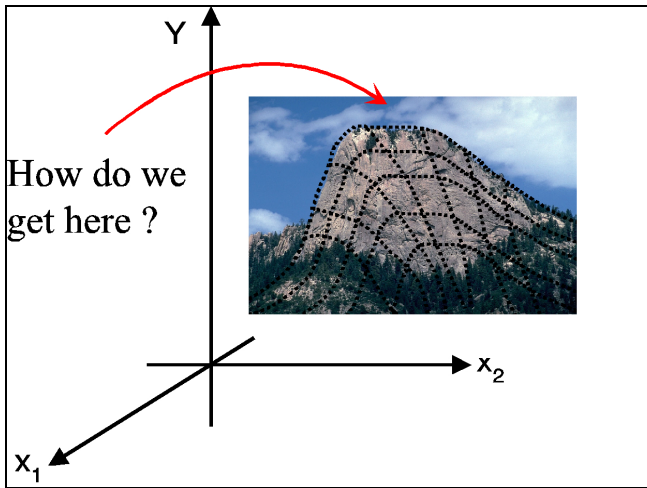
Figure 13.81: *Graphical illustration of the experimental objective* finding the optimal region.

At this stage, our intention is to describe two techniques for doing this. One technique is a graphically oriented gradient technique, and the other an automatic procedure based on running multiple simplexes in parallel. These techniques are used predominantly for extrapolation purposes, that is, for predicting outside the investigated experimental region. However, the automatic optimization procedure may be used for interpolation, as well.

Gradient techniques

Gradient techniques are useful when it is necessary to re-adjust factor settings and move the experimental design. Basically, one may envision two gradient techniques, *steepest ascent* and *steepest descent*. These are carried out using graphical tools for moving along the direction in which the best response values are expected to be encountered. Steepest ascent means that the movement is aiming for higher response values, that is, one wants to climb the mountain and reach its highest peak. Steepest descent means that the migration is directed towards lower response values, that is, one wishes to arrive at the bottom of the valley. Both gradient techniques are conveniently carried out with the help of graphical tools, e.g., response contour plots.
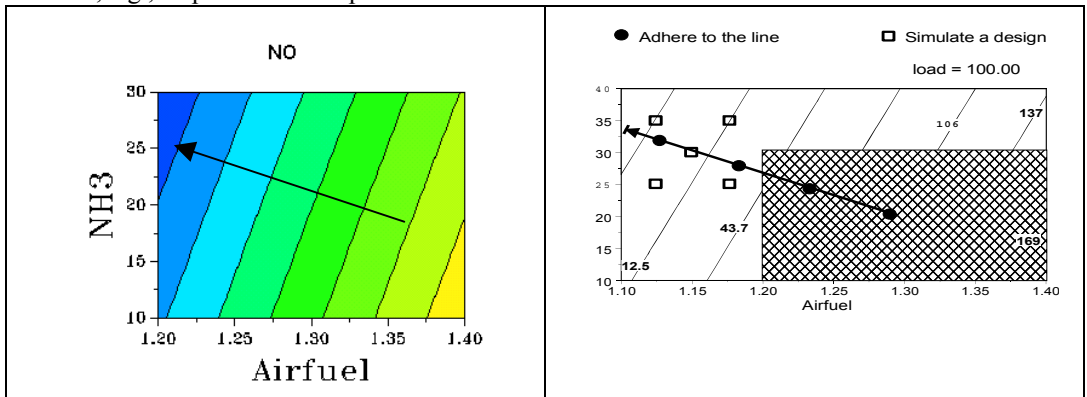
Figure 13.82: *(left) Example of steepest descent search for a lower response value.*
Figure 13.83: *(right) Steepest descent search extended outside investigated area.*

Figure 13.82 shows an example of a steepest descent application. For clarity, we skip the details of the example, and focus on the procedural steps involved. Steepest descent is implemented by first identifying an interesting direction. This direction should be perpendicular to the parallel lines of the response contour plot. This is illustrated in Figure 13.83. Here, the factor axes have been stretched out in comparison with Figure 13.82, and the domain originally mapped is depicted by the hatched area. Then one moves along the reference direction, using the existing regression model to make response predictions at equidistant points. This is symbolized by the solid circles. In fact, with this approach we use the model both for inter- and extrapolation at the same time. Alternatively, one may simulate a design, as shown by the open squares in Figure 13.83, and make predictions in a square arrangement. Regardless of the method selected, when a promising predicted point has been found, it should be tested by performing verifying experiments.

Gradient techniques work best with few responses, and when two-factor interactions are fairly small. With large two-factor interactions, response contours tend to be rather twisted, making the identification of an interesting direction difficult and ambiguous.

Gradient techniques applied to the laser welding application

Since the laser welding application comprises four factors and three responses, a large number of response contour plots are conceivable. This is an obstacle when carrying out gradient techniques. Fortunately, MODDE includes an informative 4D-response contour plot option, which permits the influence of four factors on a response to be explored at the same time. Such a plot regarding Breakage is displayed in Figure 13.84.
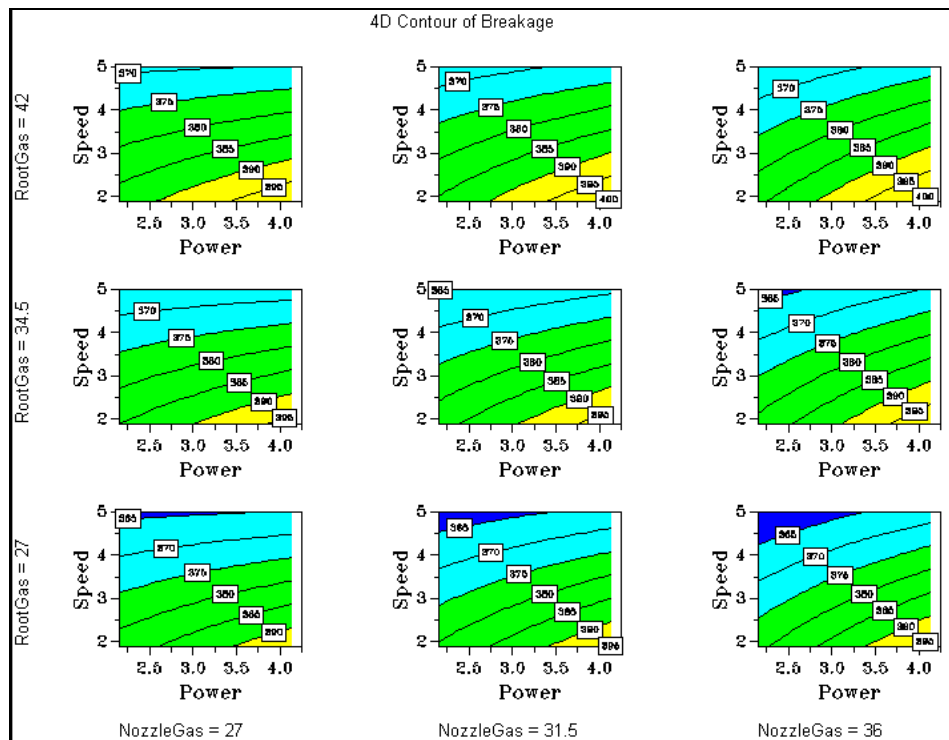
Figure 13.84: *4D-response contour plot of Breakage.*

The 3 by 3 grid of response contour plots was created by varying the four factors at three levels, low level, center level and high level. The inner factor array is made up of Power and Speed, and the outer array of NozzleGas and RootGas. Because the color coding, which goes from Breakage 360 to Breakage 410 in ten steps, is consistent over all plots, it is possible to see where the highest values of Breakage are predicted to be found. Recall that the goal of Breakage was to maximize this response, and preferably to get above 385. The predicted best point lies in the lower right-hand corner of the upper right-hand contour plot, that is, at low Speed, high Power, high NozzleGas and high RootGas.

Similar grids of contour plots can be constructed for Width and Skewness, but for reasons of brevity we do not present these. When weighing together the 4D-response contour plots of Breakage, Width and Skewness, the conclusion is that no optimal point is located within the scrutinized experimental region. We are rather close to the stipulated goals, but not sufficiently close. As a consequence, it is necessary to extrapolate. Such extrapolations can be carried out graphically with pictures such as the one shown Figure 13.84. However, a more efficient way of extrapolating is by using the automatic search routine built into MODDE. This will be discussed in the next section.

Automatic search for an optimal point

MODDE contains an optimization routine, a so called optimizer. For the optimizer to work properly, the user must specify certain initial conditions. Firstly, the roles of the factors must be set. A factor can be allowed to change its value freely in the optimization, or it may be fixed at a constant value. In addition, for the free factors, low and high variation limits must be set. Secondly, certain criteria concerning the responses must be given. A response variable may either be maximized, minimized, directed towards an interval, or excluded in the optimization. Subsequent to these definitions, the optimizer will use the obtained regression model and the performed experiments to compute eight starting points for simplexes. In principle, the optimizer then uses these simplexes together with the fitted model, to optimize desirability functions, that is, mathematical functions representing the desirabilities, or goals, of the individual responses. These goals are taken from the criteria specified for the responses.

Figure 13.85 shows one simplex in action. For simplicity, this simplex is laid out in a two-dimensional factor space. The first simplex consists of the three experiments enumerated 1 – 3. Now, the idea is to mirror the worst experiment, here #1, through the line connecting the two best experiments, here #2 and 3, and perform a new experiment, here #4, at the position where the mirrored experiment "hit" the response contour plot. This means that the new simplex consists of runs 2, 3, and 4. In the next cycle, the worst experiment is mirrored in the line connecting the two best experiments, and so on. This is repeated until the peak of the elliptical mountain is reached by one simplex. We emphasize that this is only a simplified picture of how the simplex methodology works. In reality, when a new simplex is formed through a reflection, it may be reduced or enlarged in size in comparison with the precursor simplex.
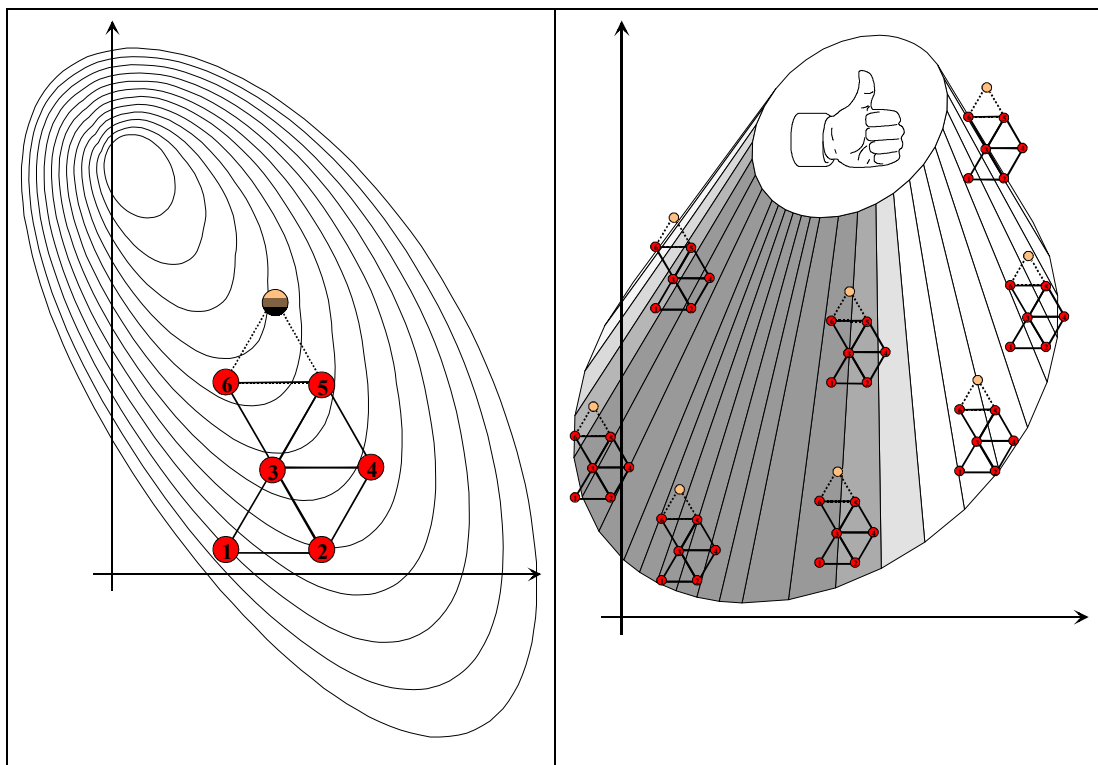
Figure 13.85: (left) An illustration of the simplex methodology.
Figure 13.86: (right) Different simplexes get different starting points in factor space.

As mentioned, the optimizer will simultaneously start up to eight simplexes, from different locations in the factor space. Eight simplexes are initiated in order to avoid being trapped at a local minimum or maximum. This is illustrated in Figure 13.86. The co-ordinates for the simplex starting points are taken from the factors with the largest regression coefficients in the regression model. At convergence, each simplex will display the obtained factor settings, and it is then possible to compute predicted response values at each point. Observe that the drawing in Figure 13.86 is a schematic valid for chasing higher response values. In reality, one could also go for a minimum, and then the relevant response function would correspond to a hole or a valley.

MODDE optimizer applied to the laser welding data

We will now relate what happened when applying the optimizer to the laser welding application. It is a good practice to conduct the optimization in stages. We recommend that the first phase is made according to the principles of interpolation, even though one may have obtained signals that extrapolation may be better. In interpolation, the low limits and high limits of the factors are taken from the factor definition. This is shown in Figure 13.87. We may disregard the block factor. It is listed, but as it has been excluded from the final model it will *not* influence the optimization.

	Factor	Role		Value	Low Limit	High Limit
1	**Power**	Free	▼		2.15	4.15
2	**Speed**	Free	▼		1.875	5
3	**NozzleGas**	Free	▼		27	36
4	**RootGas**	Free	▼		27	42
5	**$Block**	Free	▼		−1	1

Figure 13.87: Factor settings used for interpolation.

Next, it is necessary to give the desirability, or goal, of each response, that is, to specify a desired response profile. We can see in Figure 13.88 the criterion and goal set for each response. This table is understood as follows. Breakage is to maximized, and we want to obtain a value as high as 400, but will accept 385. Skewness is to be minimized, and we want to attain a value of 10, but will accept 20. Concerning Width the goal is neither maximization nor minimization, but it is desirable to direct it into an interval from 0.7 to 1.0. Technically, this is specified by setting the target to 0.85 and the limits to 0.7 and 1.0. Furthermore, it is also possible to assign weighting coefficients to each response. Such weights may range from 0.1 to 1.0. Here, we assign the same priority to all three responses, by stating that each weight should be 1.0. This weighting is an advanced topic that is not treated further in this course.

	Response	Criteria		Weight	Min	Target	Max
1	**Breakage**	Maximize	▼	1	385	400	
2	**Width**	Target	▼	1	0.7	0.85	1.0
3	**Skewness**	Minimize	▼	1		10	20

Figure 13.88: Criteria of responses used for interpolation.

First optimization - Interpolation

Now we are ready to initiate the optimization. In doing so, the optimizer will first compute pertinent starting co-ordinates, factor combinations, for launching the eight simplexes. Five starting points are derived by laying out a 2^{3-1} fractional factorial design in the three most prominent model factors, and thereafter augmenting this design with one center-point. We see in Figure 13.89 that these three factors are Power, Speed and RootGas. The fractional factorial design corresponds to the five first rows in the table. In addition, the optimizer scans through the existing worksheet and picks out those three experiments which are closest to the specified response profile. These are the three runs listed in rows 6, 7 and 8.

Iteration: 389

	1 Power	2 Speed	3 NozzleGas	4 RootGas	5 $Block	6 Breakage	7 Width	8 Skewness	9 iter	10 log(D)
1	2.15	1.875	31.5	42						
2	2.15	5	31.5	27						
3	4.15	1.875	31.5	27						
4	4.15	5	31.5	42						
5	3.15	3.4375	31.5	34.5						
6	2.15	5	27	42						
7	4.15	5	36	42						
8	2.15	5	27	27						

Figure 13.89: Factor settings for interpolation.

When pressing the arrow button, the eight simplexes will start their search for a factor combination, at which the predicted response values are as similar to the desired response profile as possible. We see the outcome of the eight simplexes in Figure 13.90. All proposed factor combinations are different, but some are rather near each other. However, none of these points is predicted to fulfill our requirements. The success of each simplex may be deduced from the numerical values of the rightmost column in Figure 13.90. This logD value represents a weighted average of the individual response desirabilities. It may be used to evaluate how the optimization proceeds. A positive logD is bad and undesirable. When logD is zero, all responses are predicted to be between their assigned target and limit values. This is a good situation. Better still is when logD is negative, and −10 is the lower, optimal, limit. Hence, in our case we are doing rather well. We are close, but not sufficiently close, to an optimal point. This means that we must try extrapolation.

Iteration: 464

	1 Power	2 Speed	3 NozzleGas	4 RootGas	5 $Block	6 Breakage	7 Width	8 Skewness	9 iter	10 log(D)
1	2.1527	2.0524	31.4824	41.9999		384.813	0.9954	24.3621	218	0.0809
2	2.1502	3.032	28.234	41.9891		380.222	0.943	19.3219	283	0.0038
3	3.8788	3.431	35.9999	41.9222		385.763	1.0956	22.1675	464	0.2132
4	4.0684	4.0663	36	41.9991		381.158	0.97	19.1584	251	0.014
5	3.5097	3.638	35.9996	41.9997		381.757	0.9872	20.4812	235	0.0525
6	2.15	3.0079	27.0006	41.3269		380.611	0.9615	19.163	339	0.0147
7	4.0683	4.014	35.9998	41.9995		381.637	0.9832	19.4143	174	0.0284
8	2.15	2.9673	27.0003	31.7101		377.567	0.9343	19.164	249	0.0586

Figure 13.90: Summary of interpolation simplex searching.

Second optimization - Extrapolation

In the extrapolation phase, it is necessary to relax the factor limits. This is usually done in several steps. One may use the results of the interpolation to get some idea of how this should be done. For instance, seven of the eight simplexes (in Figure 13.90) have RootGas close or equal to 42. This is the high setting of this factor. Hence, it is reasonable to first relax this limit. It is seen in Figure 13.91 that we decided to raise this limit to 50.

Figure 13.91: *(left) Factor settings used in first extrapolation.*
Figure 13.92: *(right) Simplex results after first extrapolation.*

Figure 13.92 shows the results of the eight new simplexes. Simplex runs #2, 4 and 7 are the best. With their factor combinations, the goals for Width and Skewness are predicted to be met, and those for Breakage are almost met. We also see that runs 4 and 7 have NozzleGas at its high limit. This made us consider relaxing the high limit of NozzleGas prior to the next round of simplex launches. As shown in Figure 13.93, the high limit of NozzleGas was changed to 45.

Figure 13.93: *(left) Factor settings used in second extrapolation.*
Figure 13.94: *(right) Simplex results after second extrapolation.*

Figure 13.94 reveals a somewhat better predicted point, that is, run #4. Interestingly, this fourth run has Power at its high level. In fact, this was also discernible for runs #4 and 7 in the first extrapolation round (Figure 13.92). Hence, we decided to increase the high level of Power from 4.15 to 5, and re-initiate the eight simplexes, the results of which are shown in Figures 13.95 and 13.96.

Figure 13.95: *(left) Factor settings used in third extrapolation.*
Figure 13.96: *(right) Simplex results after third extrapolation.*

Again the fourth simplex is predicted as the most successful, and now we are close to a factor combination at which our goals are predicted to be accomplished. Of course, we may now proceed by extrapolating even further. However, one must remember that the further the extrapolation the greater the imprecision in the predictions. Hence, at some point, the extrapolation must be terminated and the usefulness of the predicted point verified experimentally.

Local maximum or minimum

One problem with the simplex optimization approach is the risk of being trapped at a local response phenomenon, and thereby missing the global optimal point. In the optimization of a quadratic model (see Chapter 15) such a local response phenomenon might correspond to a minimum or maximum. This problem may be circumvented by taking the simplex predicted to be the best, and from its factor combination generate starting points for new simplexes. We decided to exploit this option. The best predicted point is shown as the fifth row in the table of Figure 13.97.

	1	2	3	4	5	6	7	8	9	10
	Power	Speed	NozzleGas	RootGas	$Block	Breakage	Width	Skewness	iter	log(D)
1	4.7138	4.5744	44.998	50						
2	4.7138	5	44.998	47.6459						
3	5	4.5744	44.998	47.6459						
4	5	5	44.998	50						
5	4.9988	4.8869	44.998	49.9459						

Iteration: 389

Figure 13.97: *The best predicted point is used as a starting point when generating a new series of simplexes.*

Around this point, four new starting points, rows 1-4, were defined. The co-ordinates of these new starting points were identified by moving away from the mother simplex to a distance corresponding to 20% of the factor ranges. By executing the four new simplexes and the old one, the results displayed in Figure 13.98 were obtained. Evidently, all five points are predicted to meet our experimental goals. Actually, by disregarding small variations among the decimal digits, we find that the five simplexes have converged to the same point. This means that we have not been trapped by a local response phenomenon. The identified point, with the factor settings Power = 5, Speed = 4.9, NozzleGas = 45 and RootGas = 50, is one that might be verified experimentally.

	1	2	3	4	5	6	7	8	9	10
	Power	Speed	NozzleGas	RootGas	$Block	Breakage	Width	Skewness	iter	log(D)
1	4.9619	4.8535	44.9989	49.9982		385.218	0.9893	10.2471	286	-0.2136
2	4.9998	4.8469	44.1062	49.9992		385.039	0.9842	10.7008	178	-0.2208
3	4.9869	4.8622	44.9997	49.9786		385.292	0.9904	10.1815	338	-0.2126
4	4.9995	4.8772	44.9994	50		385.225	0.9877	10.1193	193	-0.2188
5	5	4.8697	44.9703	49.9978		385.291	0.9897	10.1552	141	-0.2147

Iteration: 338

Figure 13.98: *Final results of optimization.*

Bringing optimization results into response contour plots

Prior to conducting the verifying experiment with the settings Power = 5, Speed = 4.9, NozzleGas = 45 and RootGas = 50, it is recommended that the relevance of this predicted optimal point be considered. This may be done by transferring the simplex optimization results into a graphical representation. Consider Figure 13.99. It provides a triplet of response contour plots, centered around the predicted optimal point. In the center of these plots, it is predicted that Breakage = 385, Width = 0.98 and Skewness = 10.1. The uncertainties in these predicted values are deduced from the associated 95% confidence intervals. These 95 % CIs indicate that at the optimal point Breakage is likely to vary

between 372-398, Width between 0.76-1.20, and Skewness between 5.5-18.6. The unsymmetrical confidence intervals of the latter response has to do with the fact that Skewness is modelled in its log-transformed metric, and when the predicted values are re-expressed in the original metric their associated confidence intervals become skewed.
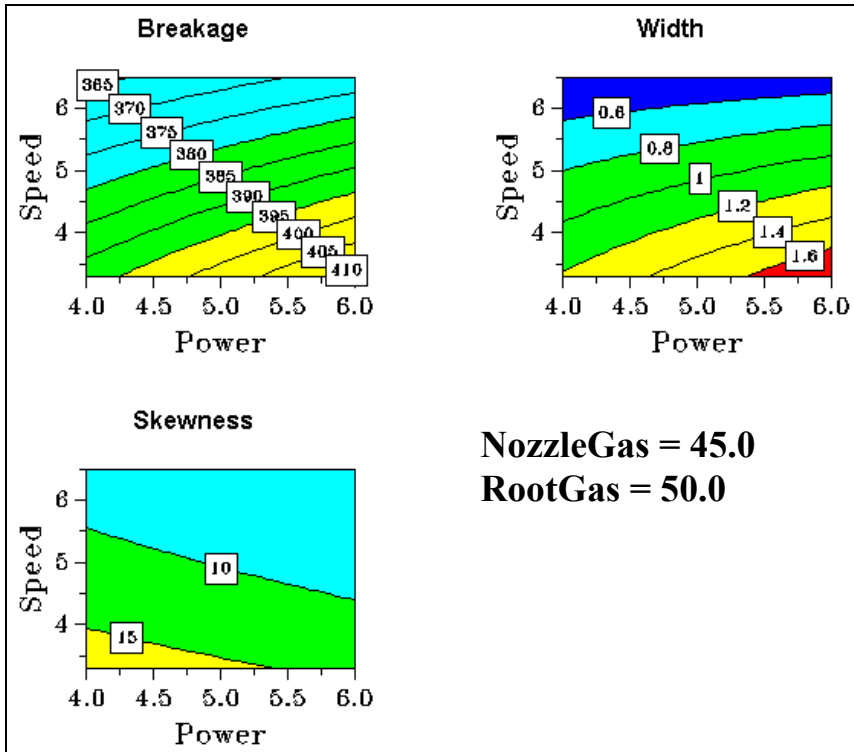
Figure 13.99: Graphical summary of the optimization results. Plots are centered around the predicted optimal point.

Summary of the laser welding application

The application of DOE in the production of gasket-free all-welded plate heat exchangers was very successful. With this approach it was possible not only to understand the mechanisms involved in the welding, but also to identify good operating conditions for the welding. We are, however, prevented from disclosing any detailed information related to the final choice of factor settings.

From an instructive point of view, the laser welding application is good for several reasons. For instance, it shows that applying DOE to technical problems often involves proceeding in stages. Initially, a screening design of only 11 experiments was laid out to probe the relevance of the defined experimental region. Once it was verified that this region was interesting, another set of 11 experiments was selected, using the fold-over technique, and used to supplement the initial experiments. With this combined set of 22 experiments it was possible to derive very good regression models for the three modelled responses Breakage, Width, and Skewness.

Furthermore, this data set also allowed us to study how tools for evaluation of raw data, tools for regression analysis, and tools for using models, were put into practice in a real situation. It should be clear from this application that the process of deriving the best possible regression model is often of an iterative nature, and frequently many repetitive modelling cycles are needed before a final model can be established. The modelling step is of crucial importance for the future outcome of a DOE project. If an investigator overlooks checking the quality of a regression model, this may have devastating consequences at later stages when the model is used for predictive purposes. Hence, strong emphasis is placed on finding the most reliable and useful model.

We also used the laser welding study for reviewing what options are available after a screening investigation has been completed. In particular, we examined the graphically-oriented gradient techniques and the more automatic search technique based on executing multiple simplexes. With the MODDE optimizer we were able to identify the factor combination Power = 5, Speed = 4.9, NozzleGas = 45 and RootGas = 50 as the predicted best point. In principle, it is then up to the experimenter to determine whether it is relevant to proceed and verify this point, or whether the optimizer should be re-used to find an alternative point at some other location in the factor space.

Our intention has been only to illustrate the procedure for identifying an optimal point. One may also discuss what such an optimal point should be used for. If the experimenter is sure that the optimum will be found in the vicinity of a predicted point, he or she may proceed directly to optimization by anchoring a composite design around the predicted optimal point. However, if he or she is not totally convinced that the optimum is within reach, it might be better to center a second screening design around the predicted best point.

Summary of fractional factorial designs

Fractional factorial designs are useful for screening large numbers of factors in few experiments. These designs are constructed by selecting fractions of corner experiments, drawn from the underlying full factorial designs. Because all possible corner experiments are not used with fractional factorial designs, these designs give rise to problems with the confounding of effects. Confounding of effects means that all effects cannot be estimated completely resolved from one another. A rapid insight into the confounding complexity of a fractional factorial design is given by its resolution. For general screening purposes, we recommend designs of resolution IV. Such designs have main effects unconfounded with two-factor interactions, which is a desirable situation. In summary, with fractional factorial designs it is feasible to answer questions, such as, which are the dominant factors? And what are their optimal ranges?

Quiz II

Please answer the following questions:

Which tools are useful for evaluating raw data?

Which tools are meaningful for diagnostic checking of a model?

What measures had to be taken to improve the initial linear model of the laser welding data?

What does interpolation mean? Extrapolation?

What options are available, after screening, when experimenting outside the experimental region is not possible or not undesirable?

What are the two primary reasons for adding complementary runs?

What is fold-over?

When and why is it necessary to log-transform responses?

What measures had to be taken to improve the interaction model of the laser welding data?

What options are available, after screening, when experimenting outside the experimental region is possible and/or desirable?

How are gradient techniques implemented?

What does simplex optimization imply?

How does the optimizer make use of existing regression models for starting simplex optimization?

What is of crucial importance regarding any final predicted best experiment?

Summary

In this chapter, we have discussed the experimental objective called screening. For this purpose, we have used the laser welding application, consisting of four factors, three responses, and two series of 11 experiments. The problem formulation steps of this application were reviewed at the beginning of the chapter. The next part of the chapter was devoted to the introduction of fractional factorial designs, and a discussion relating to their advantages and limitations was given. Much emphasis was given to a geometrical interpretation of fractional factorial designs. After the introduction of the family of fractional factorial designs, the laser welding application was outlined in detail. Both the analysis of this data-set, leading towards useful models, and what to do after screening, were addressed. At the end of the chapter, procedures for converting the screening information into concrete actions and decisions of where to carry out new experiments were considered.

14 Experimental objective: Screening (Level 3)

Objective

The objective of this chapter is to extend the treatment of theoretical concepts related to two-level fractional factorial designs. Hence, we will discuss confounding of estimated effects, and give examples of more complex confounding patterns resulting from heavily reduced fractional factorial designs. It is possible to regulate and manipulate the confounding pattern of a fractional factorial design. In order to understand how this may be accomplished, we must understand the concept of a generator. A generator is a column of alternating minus and plus signs of the computational matrix that is used to introduce a new factor. One or more generators of a design control the actual fraction of experiments that is selected and regulate the confounding pattern. With two or more generators it is beneficial to concatenate them in one single expression. This expression is called the defining relation. We will show how to derive such an expression and how to use it for overviewing a confounding pattern. Another relevant property of a fractional factorial design, its resolution, is obtainable through the defining relation. This chapter ends with a treatment of two alternative families of screening designs, the Plackett-Burman and D-optimal design families. We will highlight what they look like and when they are appropriate.

Confounding pattern

We understood in Chapter 13 that the primary incentive for using fractional factorial designs is the substantial reduction in experiments they offer. However, the price we pay for reducing the number of experiments is that our factor effects can no longer be estimated completely free of one another. The effects are said to be *confounded*, that is, to a certain degree mixed up with each other. The extent to which the effects are confounded in the laser welding design is shown by the computational matrix listed in Figure 14.1. Recall that this is a 2^{4-1} fractional factorial design.

	$x_1x_2x_3x_4$	$x_2x_3x_4$	$x_1x_3x_4$	$x_1x_2x_4$	$= x_1x_2x_3$	x_3x_4	x_2x_4	x_2x_3
	constant	x_1	x_2	x_3	x_4	x_1x_2	x_1x_3	x_1x_4
1	+	–	–	–	–	+	+	+
2	+	+	–	–	+	–	–	+
3	+	–	+	–	+	–	+	–
4	+	+	+	–	–	+	–	–
5	+	–	–	+	+	+	–	–
6	+	+	–	+	–	–	+	–
7	+	–	+	+	–	–	–	+
8	+	+	+	+	+	+	+	+

Figure 14.1: Confounding pattern of the 2^{4-1} design used in the laser welding study.

With 4 factors as many as 16 terms are theoretically conceivable, but since the design only has eight runs and hence the computational matrix only eight columns, these 16 terms can not be unequivocally determined. Here lies the disharmony, 16 terms and 8 columns, of this particular fractional factorial design. The question is then *how should 16 possible terms be spread over 8 columns*? The obvious answer is, of course, two effects per column. But, still, the most important question remains. *What is it that makes the main effects be confounded with three-factor interactions, and two-factor interactions with each other*? Will explain this in a moment. Prior to this we observe that the confounding pattern shown in Figure 14.1 is a comparatively simple situation. As long as main effects are aliased with negligible three-factor interactions, we can calculate our regression model and identify the most important factors with a high degree of certainty. This is illustrated graphically in Figure 14.2, for a regression model based on the 2^{4-1} design.
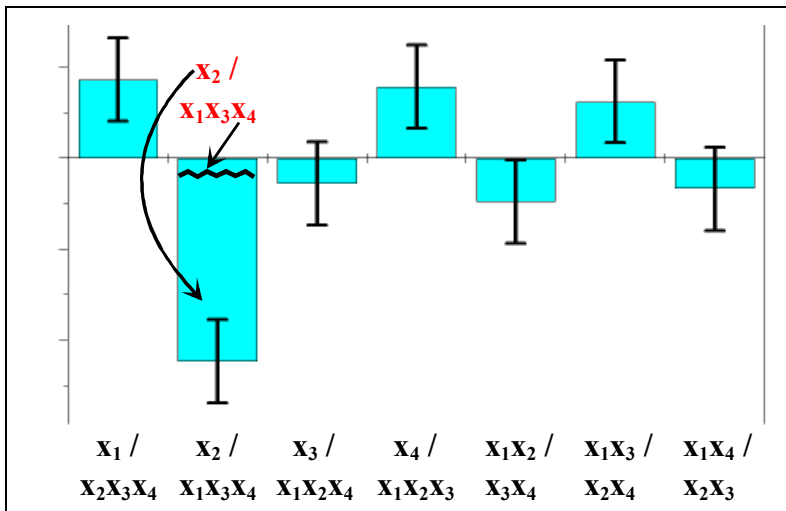
Figure 14.2: Graphical illustration of a confounding pattern of a hypothetical 2^{4-1} design. Main effects, such as x_2, are confounded with three-factor interactions, such as, $x_1x_3x_4$.

Generators

Introduction

Our aim is now to describe how it is possible to reduce the number of experiments, that is, how to select less than the full set of corner experiments. In doing so, we will start by introducing the concept of a generator. Consider the 2^3 full factorial design displayed in Figure 14.3. This is the CakeMix application. Now, let us assume that, for some reason, it is impossible to test all eight corners. The available resources for testing sanction only four corner experiments, plus an additional three replicated center-points. This means that we would like to construct the 2^{3-1} fractional factorial design in four corner experiments. Actually, as is evident from Figures 14.4 and 14.5, with this design, either of two possible half-fractions of four corner experiments can be selected.
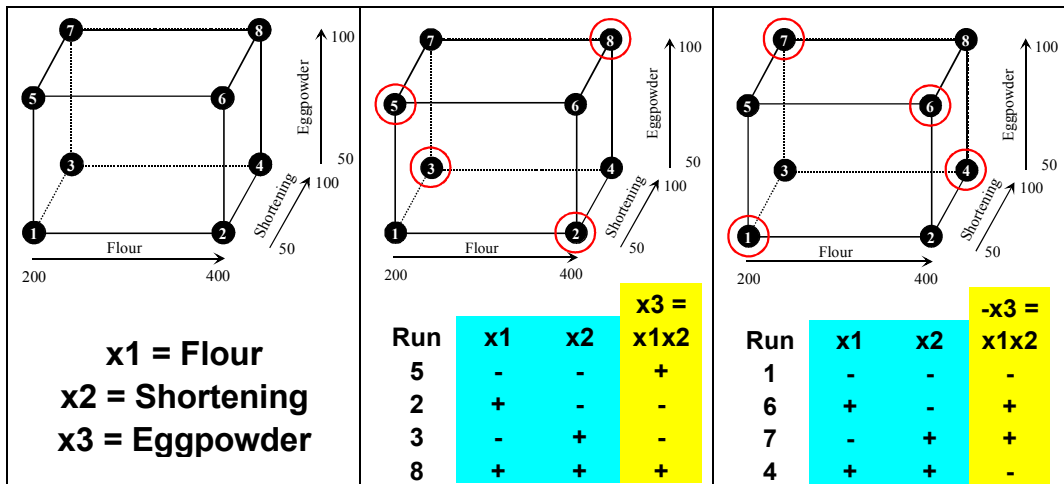
Figure 14.3: (left) Geometry of the 2^3 design used in the CakeMix application (center-points are not displayed).
Figure 14.4: (middle) Geometry of first 2^{3-1} design created with the generator $x_3 = x_1x_2$.
Figure 14.5: (right) Geometry of second 2^{3-1} design created with the generator $-x_3 = x_1x_2$.

To understand how the two half-fractions can be selected it is instructive to consider the precursor 2^2 full factorial design, which is depicted by the two columns labeled x_1 and x_2 in Figure 14.4. These columns can be used to estimate the main effects of these two factors. Attached to these two columns, we find the right-most column which may be used to estimate the x_1x_2 two-factor interaction. If we want to introduce a third factor, x_3, at the expense of some other term, the term we have to sacrifice is the two-factor interaction. It is shown in Figure 14.4 how x_3 is introduced in the design. Now we have created the first version of the 2^{3-1} fractional factorial design, and it corresponds to the four encircled experiments 2, 3, 5, and 8. It was the insertion of $x_3 = x_1x_2$ that generated this selection of four corners. Hence, the expression $x_3 = x_1x_2$ is called the *generator* of this 2^{3-1} design.

Alternatively, one may choose to incorporate x_3 into the 2^{3-1} design as $-x_3 = x_1x_2$ and obtain another selection of experiments. As seen in Figure 14.5, this alternative selection corresponds to the complementary half-fraction consisting of the runs 1, 4, 6, and 7. It was the new generator $-x_3 = x_1x_2$ that produced the complementary half fraction. From a DOE

perspective these two complementary half-fractions are equivalent, but it may well be the case that practical considerations make one half-fraction preferable to the other.

In summary, we have just shown how two different generators may be used to construct two versions of the 2^{3-1} fractional factorial design, each version encoding one half-fraction of four corner experiments. This illustrates that it is the generator that dictates which specific fraction will be selected, and thereby, indirectly, controls the confounding pattern. The 2^{3-1} fractional factorial design is ideal for illustrating the generator concept. It is not very useful in reality, however, because it results in the confounding of main effects with two-factor interactions, which in a screening design is undesirable.

Generators of the 2^{4-1} fractional factorial design

Let us now consider four factors instead of three. With four factors we may use the 2^4 full factorial design corresponding to 16 runs. These 16 factor combinations are listed in the left-hand part of Figure 14.6. Often, however, we may not want to do all these experiments, and hence it is relevant to attempt a sub-set selection. Indeed, it is possible to make a sub-set selection, and an optimal selection of experiments is then encoded by the 2^{4-1} fractional factorial design. There are two versions of the 2^{4-1} design. One version is obtained when the generator $x_4 = x_1x_2x_3$ is used. This design represents the half-fraction of experiments listed in the upper right-hand part of Figure 14.6. The other version of the 2^{4-1} design is achieved when the alternative generator $- x_4 = x_1x_2x_3$ is employed. The latter generator results in the selection of the complementary half-fraction shown in the lower right-hand part of Figure 14.6.

Run #	x_1	x_2	x_3	x_4		Run #	x_1	x_2	x_3	$x_1x_2x_3=$ x_4
1	–	–	–	–		1	–	–	–	–
2	+	–	–	–		10	+	–	–	+
3	–	+	–	–		11	–	+	–	+
4	+	+	–	–		4	+	+	–	–
5	–	–	+	–		13	–	–	+	+
6	+	–	+	–		6	+	–	+	–
7	–	+	+	–		7	–	+	+	–
8	+	+	+	–		16	+	+	+	+
							x_1	x_2	x_3	$x_1x_2x_3=$ $-x_4$
9	–	–	–	+		9	–	–	–	+
10	+	–	–	+		2	+	–	–	–
11	–	+	–	+		3	–	+	–	–
12	+	+	–	+		12	+	+	–	+
13	–	–	+	+		5	–	–	+	–
14	+	–	+	+		14	+	–	+	+
15	–	+	+	+		15	–	+	+	+
16	+	+	+	+		8	+	+	+	–

Figure 14.6: *Overview of how the 2^4 design relates to the two 2^{4-1} designs.*

In summary, we have seen that the application of generators in the 2^{3-1} and 2^{4-1} design cases is similar. In the former case, the generators are $\pm x_3 = x_1x_2$, and in the latter $\pm x_4 = x_1x_2x_3$. Thus, when only two generators exist, either will return a half-fraction of experiments.

Multiple generators

Now, we shall consider another fractional factorial design for which more than one pair of generators has to be contemplated. Consider a case with five factors. Five factors may be screened with either the 2^{5-1} or 2^{5-2} fractional factorial designs. With the former design, only two generators are defined, and hence using one of them must result in the selection of a half-fraction of 16 experiments.

On the other hand, when using the 2^{5-2} design only eight runs are encoded. When creating this design one must use the 2^3 full factorial design as starting point. The extended design matrix of the 2^3 design is shown in Figure 14.7. The best confounding pattern of the 2^{5-2} design is obtained when using the four generators $\pm x_4 = x_1 x_2$ and $\pm x_5 = x_1 x_3$. Here, the first pair of generators, $\pm x_4 = x_1 x_2$, means that the fourth factor is introduced in the $x_1 x_2$ column. Similarly, the second pair of generators, $\pm x_5 = x_1 x_3$, implies that the fifth factor is inserted in the $x_1 x_3$ column. It is easily realized that with two pairs of generators, four combinations of them are possible, i.e., the forth and fifth factors may be introduced in the design as $+x_4/+x_5$, $+x_4/-x_5$, $-x_4/+x_5$, or $-x_4/-x_5$. Each such combination prescribes the selection of a unique quarter-fraction comprising eight experiments. From a DOE perspective, these four quarter-fractions are equivalent, but the experimentalist may prefer one of them to the others.

	x_1	x_2	x_3	$\pm x_4 =$ $x_1 x_2$	$\pm x_5 =$ $x_1 x_3$	$x_2 x_3$	$x_1 x_2 x_3$
1	–	–	–	+	+	+	–
2	+	–	–	–	–	+	+
3	–	+	–	–	+	–	+
4	+	+	–	+	–	–	–
5	–	–	+	+	–	–	+
6	+	–	+	–	+	–	–
7	–	+	+	–	–	+	–
8	+	+	+	+	+	+	+

Figure 14.7: *How the fourth and fifth factors may be introduced in a precursor 2^3 design.*

Defining relation

Introduction

So far, we have mainly confronted fractional factorial designs dealing with three, four, or five factors, with comparatively few generators. However, when screening many more factors, one has to work with designs founded on several generators. Since it is not very practical to keep track of many generators in terms of isolated generator expressions, it would seem appropriate to replace such isolated expressions with some single relation tying them all together. Indeed, this is possible with what is called the *defining relation* of a design. The defining relation of a design is a formula derived from all its generators, that allows the calculation of the confounding pattern. Will we illustrate this using the 2^{4-1} fractional factorial design.

Rules:

(1) $X_1*I=X_1$ (2) $X_1*X_1=X_1^2=I$

X_1	$*$	I	$=$	X_1		X_1	$*$	X_1	$=$	I
-		+		-		-		-		+
+		+		+		+		+		+
-		+		-		-		-		+
+		+		+		+		+		+

Step 1:   Identify generator(s): $X_4=X_1X_2X_3$
Step 2:   Multiply both sides by X_4: $X_4^2=X_1X_2X_3X_4$
Step 3:   Apply rule 2 $I=X_1X_2X_3X_4$
This is the defining relation for the 2^{4-1} design
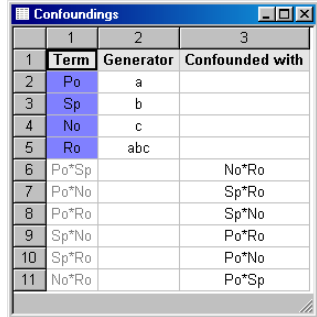
Figure 14.8: How to derive the defining relation of the 2^{4-1} fractional factorial design.

In order to derive the defining relation of the 2^{4-1} design, we must first understand two computation rules pertaining to column-wise multiplication of individual column elements. This is illustrated in Figure 14.8.

The first rule says that when multiplying the elements in any column, say column x_1, by the elements of the identity column, I, one obtains the same column x_1. The second rule says that when multiplying the elements in any column by a column of identical elements, one gets I. We will now put these rules into operation. The first step in the derivation of the defining relation is to identify the generator(s). In the case of the 2^{4-1} design used in the laser welding application, the generator is $x_4 = x_1x_2x_3$. Then, in the second step, both sides of this equality is multiplied by x_4. Finally, in step 3, upon applying rule 2, it is seen that $I = x_1x_2x_3x_4$. This is the defining relation of the 2^{4-1} fractional factorial design.

Use of defining relation

The defining relation is very useful in that it immediately yields the confounding pattern. Figure 14.9 shows how to use the defining relation of the 2^{4-1} design, that is, the design underlying the first series of experiments in the laser welding case. To compute the confounding for a given term, one may proceed as follows: Step 1: Identify the defining relation; Step 2: Identify the term of interest, for instance x_1, and multiply both sides of the defining relation by that term; Step 3: Apply rule 1 to the left-hand side and rule 2 to the right-hand side; Step 4: Apply rule 1 to the right-hand side. By computing the confoundings of all other terms in this fashion, we get the results summarized in Figure 14.9. For comparative purposes, we have shown in Figure 14.10 how MODDE lists the confounding pattern of the initial laser welding design. For clarity, three-factor interactions have been omitted, as they are assumed to be of limited utility.

Defining relation	$I = x_1x_2x_3x_4$	
What is x_1 confounded with?	$x_1 = x_2x_3x_4$	
Step 1: $I = x_1x_2x_3x_4$		
Step 2: $x_1I = x_1^2x_2x_3x_4$		
Step 3: $x_1 = Ix_2x_3x_4$		
Step 4: $x_1 = x_2x_3x_4$		
What is x_2 confounded with?	$x_2 = x_1x_3x_4$	
What is x_2 confounded with?	$x_3 = x_1x_2x_4$	
What is x_4 confounded with?	$x_4 = x_1x_2x_3$	
What is x_1x_2 confounded with?	$x_1x_2 = x_3x_4$	
What is x_1x_3 confounded with?	$x_1x_3 = x_2x_4$	
What is x_1x_4 confounded with?	$x_1x_4 = x_2x_3$	
What is x_2x_3 confounded with?	$x_2x_3 = x_1x_4$	
What is x_2x_4 confounded with?	$x_2x_4 = x_1x_3$	
What is x_3x_4 confounded with?	$x_3x_4 = x_1x_2$	

Confoundings

	Term	Generator	Confounded with
2	Po	a	
3	Sp	b	
4	No	c	
5	Ro	abc	
6	Po*Sp		No*Ro
7	Po*No		Sp*Ro
8	Po*Ro		Sp*No
9	Sp*No		Po*Ro
10	Sp*Ro		Po*No
11	No*Ro		Po*Sp

Figure 14.9: (left) How to compute the confounding for each individual term estimable with the initial laser welding model.

Figure 14.10: (right) An overview of the confounding pattern of the 2^{4-1} design (laser welding data).

Defining relation of the 2^{5-2} fractional factorial design

The defining relation of the 2^{4-1} fractional factorial design is comparatively simple, and is not the most useful example for demonstrating the utility of this kind of expression. The key importance of the defining relation is better appreciated when considering the 2^{5-2} design. Recall that for this design two pairs of generators, $\pm x_4 = x_1x_2$ and $\pm x_5 = x_1x_3$, apply. By processing these generators according to the procedure previously outlined, it is possible to derive the defining relation $I = x_1x_2x_4 = x_1x_3x_5 = x_2x_3x_4x_5$, which is shown in Figure 14.11. In this expression, the last term, or "word", was obtained through multiplication of the generator-pairs. As seen in Figure 14.11, this defining relation may now be used to derive the confounding pattern of the 2^{5-2} fractional factorial design. Figure 14.11 lists the complete confounding pattern and Figure 14.12 a somewhat simplified confounding pattern as it is rendered in MODDE. We emphasize that in reality the experimenter does not have to pay a great deal of attention to the confounding pattern or the defining relation, as this is automatically taken care of by the software.

$$I = x_1x_2x_4 = x_1x_3x_5 = x_2x_3x_4x_5$$

$$x_1 = x_2x_4 = x_3x_5 = x_1x_2x_3x_4x_5$$
$$x_2 = x_1x_4 = x_1x_2x_3x_5 = x_3x_4x_5$$
$$x_3 = x_1x_2x_3x_4 = x_1x_5 = x_2x_4x_5$$
$$x_4 = x_1x_2 = x_1x_3x_4x_5 = x_2x_3x_5$$
$$x_5 = x_1x_2x_4x_5 = x_1x_3 = x_2x_3x_4$$
$$x_1x_2 = x_4 = x_2x_3x_5 = x_1x_3x_4x_5$$
$$x_1x_3 = x_2x_3x_4 = x_5 = x_1x_2x_4x_5$$
$$x_1x_4 = x_2 = x_3x_4x_5 = x_1x_2x_3x_5$$
$$x_1x_5 = x_2x_4x_5 = x_3 = x_1x_2x_3x_4$$
$$x_2x_3 = x_1x_3x_4 = x_1x_2x_5 = x_4x_5$$
$$x_2x_4 = x_1 = x_1x_2x_3x_4x_5 = x_3x_5$$
$$x_2x_5 = x_1x_4x_5 = x_1x_2x_3 = x_3x_4$$
$$x_3x_4 = x_1x_2x_3 = x_1x_4x_5 = x_2x_5$$
$$x_3x_5 = x_1x_2x_3x_4x_5 = x_1 = x_2x_4$$
$$x_4x_5 = x_1x_2x_5 = x_1x_3x_4 = x_2x_3$$

Confoundings

	1	2	3	4
1	Term	Generator	Confounded with	Confounded with
2	x1	a	x2*x4	x3*x5
3	x2	b	x1*x4	
4	x3	c	x1*x5	
5	x4	ab	x1*x2	
6	x5	ac	x1*x3	
7	x1*x2		x4	
8	x1*x3		x5	
9	x1*x4		x2	
10	x1*x5		x3	
11	x2*x3		x4*x5	
12	x2*x4		x1	x3*x5
13	x2*x5		x3*x4	
14	x3*x4		x2*x5	
15	x3*x5		x1	x2*x4
16	x4*x5		x2*x3	

Figure 14.11: (left) The complete confounding pattern of the 2^{5-2} design.
Figure 14.12: (right) A simplified confounding pattern of the 2^{5-2} design, as given in MODDE.

Resolution

The defining relation is useful in that it directly gives the confounding pattern of a fractional factorial design. In addition, this expression is of central importance when it comes to uncovering the *resolution* of a design. The concept of design resolution has to do with the complexity of the confounding pattern. A design of high resolution conveys factor estimates with little perturbation from other effect estimates. On the other hand, a design of low resolution experiences complicated confounding of effects.

In order to understand the resolution of a design, we need to consider the defining relation of the 2^{5-2} design. We remember that the relevant defining relation was $I = x_1x_2x_4 = x_1x_3x_5 = x_2x_3x_4x_5$. This expression is said to consist of three words. Two of the words have the length three and one the length four. The key point is that the resolution of a design is defined as the *length of the shortest word in the defining relation*. Because the shortest word in the defining relation of the 2^{5-2} design has length 3, this design is said to be of resolution III.

The vast majority of screening designs used have resolutions in the range of III to V. We now give a summary of their confounding properties. In a resolution III design, main effects are confounded with two-factor interactions, and two-factor interactions with each other. This is a complex confounding pattern that should be avoided in screening. However, resolution III designs are useful for robustness testing purposes (see Chapter 17). In contrast to this, the confounding pattern of a resolution IV design is more tractable in screening. Here, main effects are unconfounded with two-factor interactions, but the two-factor interactions are still confounded with each other. With a resolution IV design the experimentalist has access to a design with an appropriate balance between the number of factors and the number of experiments. Resolution IV designs are recommended for screening. Further, with a design of resolution V, even the two-factor interactions

themselves are unconfounded. This means that resolution V designs are almost as good as full factorial designs. Thus, although a resolution V design would be a technically correct choice for screening, it requires unnecessarily many runs.

Summary of fractional factorial designs

The topics discussed so far in this chapter – confounding pattern, generator, defining relation, and resolution of fractional factorial designs – are intimately connected (see Figure 14.13). It has been our intention to provide an account of these fundamental design concepts. In the design generation phase, everything starts with the selection of one or more generators. The chosen generators control which fraction of experiments is to be investigated, and also regulate the confounding pattern. When several generators are in operation these may be concatenated through the defining relation. The major advantage of the defining relation is that it is an expeditious route towards an understanding of the confounding pattern. Since the complexity of a confounding pattern may vary considerably, it is desirable to somehow quantify the resolution of a design. This has been agreed upon by defining that the resolution of a fractional factorial design equals the length of the shortest word in the defining relation.
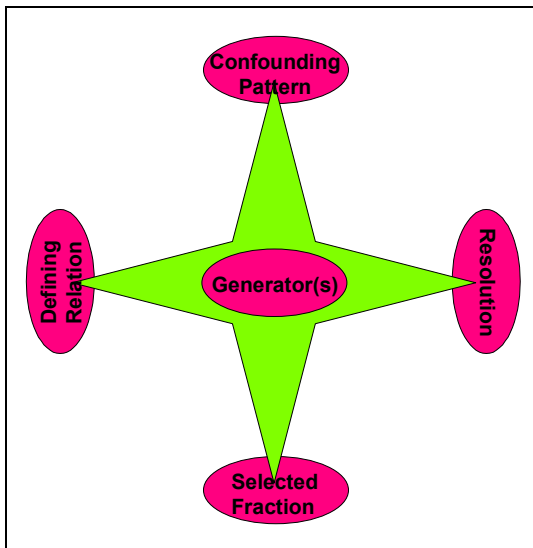
Figure 14.13: *The selected generators control the confounding pattern and the selected fraction of experiments. Indirectly, this means that the selected generators also influence the shape of the defining relation and the resolution of the design.*

In summary, we wish the reader to be familiar with these concepts and to have a basic understanding of how they work. In a real situation, any standard DOE software will take care of the technical details, and really the only thing an experimenter should worry about is to make sure that a resolution IV design is selected in screening. Occasionally, there may be no resolution IV design available, and then a resolution V design should be identified. This concludes the discussion regarding the fractional factorial design family. The remaining part of this chapter will deal with two alternative screening design families, those entitled Plackett-Burman and D-optimal designs.

Plackett-Burman designs

We will now acquaint ourselves with the family of designs called Plackett-Burman (PB). We start this introduction by addressing when these designs are useful. The primary rationale behind PB-designs is that they fill certain gaps occurring in the number of experiments encoded by fractional factorial designs. Consider the table shown in Figure 14.14, which shows the recommended use of PB-designs. This table is organized as follows: The first and leftmost column indicates the recommended number of factors to investigate, the second column the maximum number of factors that can be mapped, and the third column the number of necessary corner-experiments (runs). The last column is a remark column.

Since the two-level fractional factorial designs are only available when the number of encoded runs is a power of 2, there will be certain jumps in the required number of runs. The commonly used fractional factorial designs have 8, 16, or 32 runs. In some cases these numbers are incompatible with the experimental reality. In other cases it might be felt that the transition from 8 to 16, or 16 to 32 runs is intolerable, that is, these gaps are too wide. Fortunately, some of this strain can be relieved by considering the Plackett-Burman design family.

Rec. # factors	Max. # factors	# runs	Remark
5	7	8	Use Frac Fac
9	11	**12**	
13	15	16	Use Frac Fac
17	19	**20**	
21	23	**24**	
25	27	**28**	
29	31	32	Use Frac Fac

Figure 14.14: *Overview of when PB designs are useful.*

The Plackett-Burman designs are orthogonal two-level screening designs, in which the number of runs is a multiple of 4. This means that the PB-designs very neatly close the gap between the various fractional factorial designs. It is seen in Figure 14.14 that the PB-designs of 12, 20, 24 and 28 runs are particularly relevant. However, whenever 8, 16 or 32 runs are acceptable, fractional factorial designs are recommended. Finally, in this table we give a recommended number of factors. This is because using somewhat fewer factors than the theoretical maximum ensures that some degrees of freedom are available in the data analysis. Also observe that three replicated center-points should always be added to each design.

Properties of PB-designs

We will now look at some properties of PB-designs, by inspecting the 12 run PB-design in 11 factors displayed in Figure 14.15. The tricky part in the construction of a PB-design lies in obtaining an appropriate starting row of alternating minus one and plus one digits. Fortunately, we need not bother with this problem thanks to the pioneering work of Plackett, Burman, and others. We simply use their findings and this is also built into MODDE.

	1	5	6	7	8	9	10	11	12	13	14	15
	Exp No	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11
1	1	1	-1	1	-1	-1	-1	1	1	1	-1	1
2	2	1	1	-1	1	-1	-1	-1	1	1	1	-1
3	3	-1	1	1	-1	1	-1	-1	-1	1	1	1
4	4	1	-1	1	1	-1	1	-1	-1	-1	1	1
5	5	1	1	-1	1	1	-1	1	-1	-1	-1	1
6	6	1	1	1	-1	1	1	-1	1	-1	-1	-1
7	7	-1	1	1	1	-1	1	1	-1	1	-1	-1
8	8	-1	-1	1	1	1	-1	1	1	-1	1	-1
9	9	-1	-1	-1	1	1	1	-1	1	1	-1	1
10	10	1	-1	-1	-1	1	1	1	-1	1	1	-1
11	11	-1	1	-1	-1	-1	1	1	1	-1	1	1
12	12	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
13	13	0	0	0	0	0	0	0	0	0	0	0
14	14	0	0	0	0	0	0	0	0	0	0	0
15	15	0	0	0	0	0	0	0	0	0	0	0

Figure 14.15: A 12 run Plackett-Burman design in 11 factors. Note that the worksheet contains 3 appended replicates.

Once the first row has been defined the construction of a Plackett-Burman design is simple. It goes as follows. The first row is permuted cyclically one-step to the right and shifted one-step down. This means that the entries of the first row will appear in the same order in the second row, but be shifted to the right. Next, the third row is generated by shifting the second row one-step to the right, and so on. This continues until the next shifting would have resulted in the initial row, whereupon the permutation is terminated. In the current case, this means that row 11 is the last one to be created through the permutation procedure. Finally, a row of only minus one figures is added in the bottom of the design, followed by the replicated center-points (Figure 14.15). This 12 run PB-design, supplemented with three center-points, well exemplifies the compactness of PB-designs. They require very few experiments per investigated factor. Actually, PB-designs are special fractional factorial designs of resolution III, and hence they can only be used to estimate main effects. PB-designs are very useful in robustness testing. With PB-designs it is possible to simultaneously test both quantitative and qualitative factors.

D-optimal designs

D-optimal designs are computer generated designs that are valuable alternatives to fractional factorial and PB-designs whenever the experimenter wants to create a non-standard design. The theory concerning the construction of D-optimal designs is an advanced topic, which is dealt with in Chapter 18. Here we focus on the practical aspects of applying D-optimal design. We envision six screening situations, illustrated in Figures 14.16 – 14.21, in which D-optimal designs are particularly useful:

Irregular experimental region (Figure 14.16): This occurs when there is a part of the experimental region in which we are unable or unwilling to do experiments. Here, classical designs of regular geometry are less applicable, and an adaptable D-optimal design of irregular geometry is more appropriate.

Mixture design with irregular region (Figure 14.17): This usually occurs when there are lower and upper bounds, other than 0 and 1, of the mixture factors. Here classical mixture designs of regular geometry are inapplicable.

Qualitative factors at many levels (Figure 14.18): With qualitative factors at many levels factorial designs tend to encode a large number of runs. With a D-optimal design it is possible to obtain a balanced spread of the experiments, yet keeping the number of runs low.

Resources for a fixed number of runs (Figure 14.19): Assume that we have starting material for only 14 experiments and do not want to do either a fractional factorial design with 8+3 runs, or a PB-design with 12+3 runs, rather we want a design with *exactly* 14 experiments. This is solved with a D-optimal design.

Fitting of a special regression model (Figure 14.20): Applicable when, for instance, an interaction model in three factors must be updated with a single quadratic term, but there is no interest in upgrading to the full quadratic model.

Inclusion of existing experiments (Figure 14.21)*:* Occurs when we have a limited number of interesting experiments already done, which cover a small region, and want to supplement these in the best possible manner.
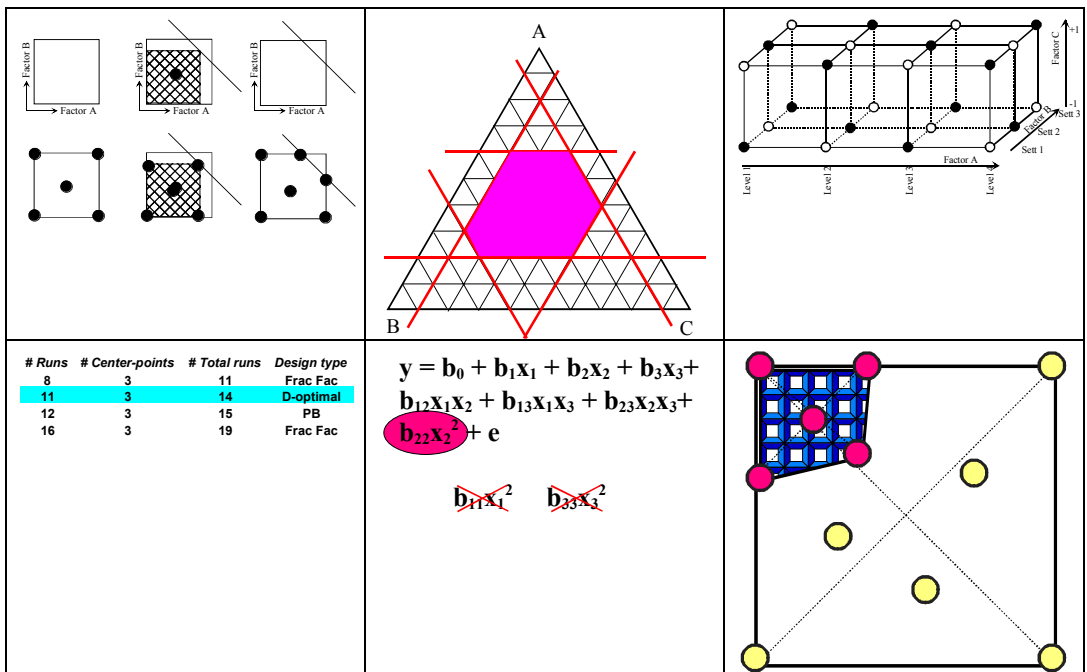
Figure 14.16: (upper left) When to use D-optimal design: Irregular experimental region.
Figure 14.17: (upper middle) When to use D-optimal design: Mixture design of irregular region.
Figure 14.18: (upper right) When to use D-optimal design: Qualitative factors at many levels.
Figure 14.19: (lower left) When to use D-optimal design: Resources for fixed number of runs.
Figure 14.20: (lower middle) When to use D-optimal design: Fitting of special regression model.
Figure 14.21: (lower right) When to use D-optimal design: Inclusion of existing experiments.

In summary, a D-optimal design is useful in screening when we want to create a non-standard design. Observe that such non-standard designs often exhibit irregular geometric properties and hence might be a little more difficult to analyze than standard DOE-protocols.

Quiz

Please answer the following questions:

What is the confounding pattern of a fractional factorial design?

What is a generator?

Which important features of a fractional factorial design are controlled by the generators?

What is a defining relation?

What is the resolution of a fractional factorial design?

Which resolution is recommended for screening?

Which resolution is recommended for robustness testing?

What is a Plackett-Burman (PB) design?

When are PB-designs of interest?

Which kind of effects may be estimated with PB-designs?

What are D-optimal designs?

When is D-optimal design applicable in screening?

Summary

In this chapter, we have outlined details of fractional factorial designs. In principle, all the important features of fractional factorial designs are controlled by the selected generator(s). The features we have concentrated on are: the selected fraction of experiments, the confounding pattern, the defining relation and the resolution. Fortunately, with MODDE, the user only has to make sure that a resolution IV design is selected, and the rest will be automatically taken care of. Furthermore, we introduced the Plackett-Burman and D-optimal design families, and pointed out situations where these may be more useful than the fractional factorial designs.

15 Experimental objective: Optimization (Level 2)

Objective

The objective of this chapter is to use the DOE framework presented in Chapters 1-12 and discuss an optimization study. We will describe the truck engine application, in which the goal is to adjust three controllable factors so that a favorable response profile with regard to fuel consumption and exhaust gases is obtained. The experimental protocol underlying the truck engine application originates from the composite design family. We will provide a detailed account of two commonly used members of the composite family, the designs entitled CCC and CCF. These abbreviations refer to central composite circumscribed and central composite face-centered. Further, the analysis of the engine data involves the repeated use of a number of tools for raw data evaluation and regression analysis, which were introduced in early chapters. In addition, much emphasis will be given to how to solve an optimization problem using simplex searching and graphical representation of results.

Background to General Example 2

At Volvo Car Corporation and Volvo Truck Corporation, DOE is commonly applied in the early stages of new engine developments (Volvo Technology Report, #2, 1997). In this context, common goals are to reduce fuel consumption and reduce levels of unwanted species in exhaust gases. One way of decreasing the emission of nitrous oxides (NO_x) from a combustion engine is to decrease the combustion temperature. By lowering this temperature, smaller amounts of NO_x are formed by reaction between atmospheric nitrogen and oxygen. It is possible to lower this temperature by supplying an inert gas, which does not react but only absorbs heat. This kind of gas may originate from the car exhaust, and hence the term exhaust gas recirculation (EGR) is often used. A negative consequence of the EGR principle is that less air is available for combustion, since the EGR mass does not allow as much air to enter the cylinder. This may, in turn, cause elevated levels of Soot in the exhaust gases. By manipulating, for instance, inlet cylinder pressure, inlet temperature, and timing of fuel injection with regard to piston movement ("needle lift"), it is possible to influence the levels of Soot and NOx in the exhausts. A schematic of a typical testing environment is given in Figure 15.1.
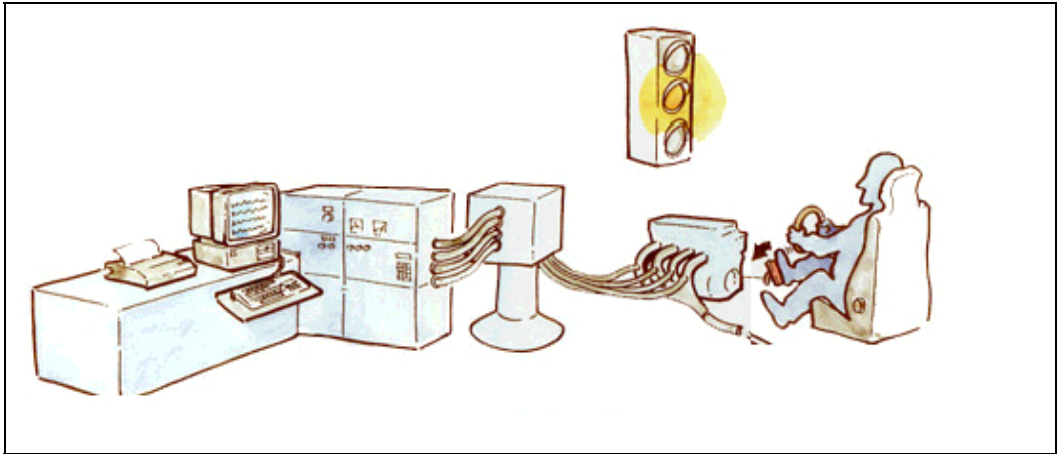
Figure 15.1: *A schematic of a testing environment used in optimization of combustion engines.*

The example that we will study was kindly contributed by Sven Ahlinder at Volvo Technical Development (VTD). It deals with the optimization of a truck piston engine. Sven Ahlinder studied how three factors, that is, *air mass used in combustion*, *exhaust gas re-circulation*, and *timing of needle-lift*, affected three responses, that is, *fuel consumption*, and levels of *NOx* and *Soot* in the exhaust gases. The units and the high and low settings of these factors are given in Figure 15.2. Analogously, Figure 15.3 discloses the units of the three responses.

Figure 15.2: *(left) Overview of factors studied in the truck engine application.*
Figure 15.3: *(right) Overview of measured responses in the truck engine application.*

As far as the responses are concerned, the desired response profile was fuel consumption as low as possible with a target of 230 mg/stroke (of piston), and simultaneously low levels of NOx and Soot in the exhaust emissions with targets of 25 mg/s and 0.5 mg/s, respectively.

This application is typical in that only a limited number of factors are studied. In optimization, usually between 2 and 5 factors are explored. This means that one can afford to make comparatively many experiments per varied factor and thus acquire a detailed understanding of the relationships between factors and responses. Sven Ahlinder used a CCF design in 17 experiments to determine whether the stipulated response profile was attainable. This design corresponds to a classical *response surface modelling*, RSM, design. It is extensively used in optimization as it supports quadratic polynomial models.

Problem Formulation (PF)

Selection of experimental objective

As discussed in Chapter 4, the problem formulation (PF) phase is important in DOE. In the truck engine application, the selected experimental objective was *optimization*. In optimization, the important factors, usually between 2 and 5, have already been identified, and we now want to extract detailed information about them. It is of interest to reveal the nature of the relationships between these few factors and the measured responses. For some factors and responses the relationships might be linear, for others non-linear, that is, curved. For some factors and responses there exists a positive correlation, for others a negative correlation. These relationships are conveniently investigated by fitting a quadratic regression model.

A quadratic model may be used to predict response values for any factor combination in the region of interest. It may also be used for optimization, that is, for identifying the factor setting(s) fulfilling the desired response profile. Sometimes only *one* distinct point is predicted, sometimes several geometrically close factor combinations might be plausible. The latter case then corresponds to a so called *region of operability*. It is also possible that points clearly separated in factor space may be predicted to share the same response profile. We then have a situation of multiple optima.

Specification of factors

The second step of the problem formulation concerns the specification of factors. This is quite simple in optimization, because at this stage the important factors have already been identified. Since these are usually limited in number, it is not necessary to use the Ishikawa diagram for system overview. This diagram, which was introduced in Chapter 13, is most useful in screening. In the truck engine application, three factors were varied. Figures 15.4 – 15.6 show their details. The first factor, air mass used in combustion, is varied between 240 and 284 kg/h. The second factor, exhaust gas-recirculation, is varied between 6 and 12%. The last factor, NeedleLift, which has to do with the *timing* of the ignition procedure, is expressed in degrees before top dead-center (°BTDC) and varies between –5.78 and 0. The appropriateness of these values were determined during a previous screening phase. Notice, that all three factors are controlled and quantitative.
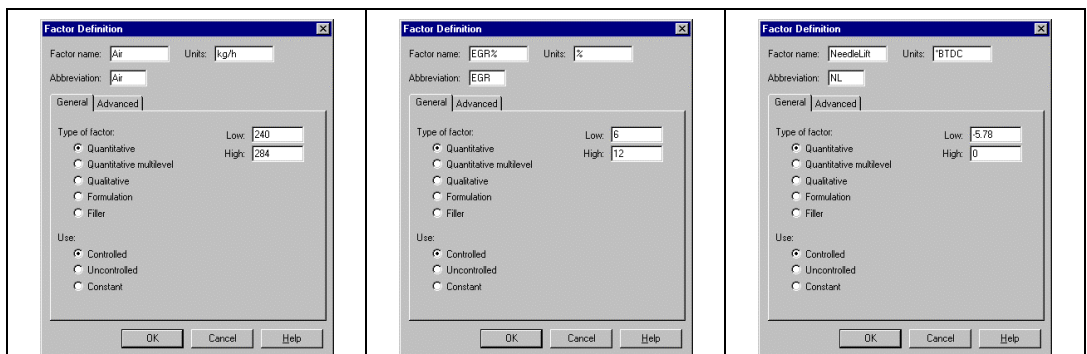
Figure 15.4: (left) Factor definition of air mass used in combustion.
Figure 15.5: (middle) Factor definition of exhaust gas-recirculation.
Figure 15.6: (right) Factor definition of timing of needle lift.

Specification of responses

The next step in the problem formulation deals with the specification of the responses. It is important to select responses that are relevant to the experimental goals. In the truck engine application, low fuel consumption and low levels of certain species in the exhaust emissions were the important properties. The investigator Sven Ahlinder registered four responses, but for simplicity we will study only three of them, namely the fuel consumption, and the amounts of NOx and Soot in the exhaust emissions. All details regarding these responses are given in Figures 15.7 – 15.9.

Figure 15.7: *(left) Response definition of fuel consumption.*
Figure 15.8: *(middle) Response definition of exhaust gas-recirculation.*
Figure 15.9: *(right) Response definition of timing of needle lift.*

The three responses are quantitative and untransformed. The first response is expressed in mg/st, the second in mg/s, and the third in mg/s. The overall goal is a response profile with a low value of fuel with a target of 230 mg/stroke, a low value of NOx with a target of 25 mg/s, and a low value of Soot with 0.5 mg/s specified as the target.

Selection of regression model

The selection of an appropriate regression model is the fourth step of the problem formulation. Remember that we may select between three different types of regression model (linear, interaction, quadratic). A quadratic model in all factors is a sound choice for the optimization objective. The reason for this is that the final goal is optimization and this often involves modelling curved response functions. Indeed, a quadratic model was selected in the truck engine application.

Quadratic models are flexible and can mimic many different types of response function. Figures 15.10 – 15.13 illustrate that quadratic models may take the shapes of stationary ridges, rising ridges, saddle surfaces, or mountains. This explains the popularity of quadratic models. Compared to a linear or an interaction model, however, the quadratic model requires many experiments per varied factor. Ideally each factor should be explored at five levels to allow a reliable quadratic model to be postulated. But, there are also good optimization designs available, which utilize only three levels per explored factor.
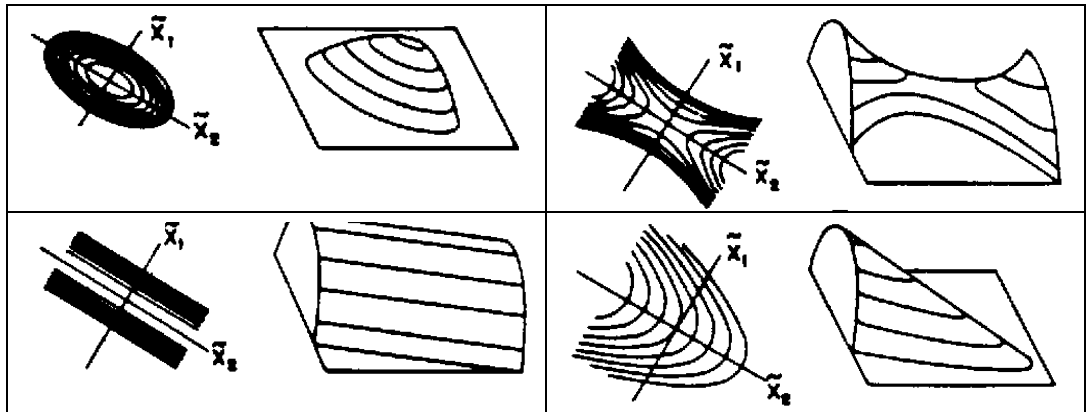
Figure 15.10: (upper left) Surfaces mimicked by quadratic models: Top of mountain.
Figure 15.11: (upper right) Surfaces mimicked by quadratic models: Saddle surface.
Figure 15.12: (lower left) Surfaces mimicked by quadratic models: Stationary ridge.
Figure 15.13: (lower right) Surfaces mimicked by quadratic models: Rising ridge.

Generation of design and creation of worksheet

The last two stages of the problem formulation deal with the generation of the statistical experimental design and the creation of the corresponding worksheet. For the truck engine application we will treat these two steps together. The experimenter selected a *central composite face-centered*, CCF, design in 17 runs. This is a standard design, which supports a quadratic model. Its worksheet is given in Figure 15.14.

	Exp No	Exp Name	Run Order	Incl/Excl	Air	EGR%	NeedleLift	Fuel	NOx	Soot
1	1	N1	2	Incl	241	6.1	−5.78	236.637	25.3414	0.14
2	2	N2	4	Incl	284	6	−5.78	254.722	26.2859	0.06
3	3	N3	10	Incl	241	12	−5.78	242.841	11.7348	1.15
4	4	N4	5	Incl	283	12.1	−5.78	262.626	12.6292	0.67
5	5	N5	14	Incl	241	6.2	0	217.069	29.6733	0.45
6	6	N6	11	Incl	283	6.1	0	230.527	32.7326	0.22
7	7	N7	6	Incl	242	12.1	0	221.274	13.6371	3.3
8	8	N8	13	Incl	284	11.8	0	236.25	15.2	1.94
9	9	N9	15	Incl	241	9	−3.16	231.77	17.48	1.24
10	10	N10	16	Incl	284	9.2	−3.16	248.064	19.3567	0.47
11	11	N11	3	Incl	261	6	−3.16	242.616	28.0458	0.19
12	12	N12	12	Incl	261	12	−3.16	245.303	12.7836	1.29
13	13	N13	7	Incl	262	9.3	−5.78	256.6	17.6	0.26
14	14	N14	9	Incl	262	9	0	227.266	22.0432	1.61
15	15	N15	17	Incl	261	9.2	−3.16	247.479	19.0623	0.44
16	16	N16	1	Incl	261	9	−3.16	243.15	19.125	0.35
17	17	N17	8	Incl	261	9.1	−3.16	242.982	19.1812	0.47

Figure 15.14: The worksheet of the truck engine application.

The three columns of the worksheet numbered 5, 6, and 7 represent the identified design. In the three right-most columns of the worksheet, we can see the measured response values.

The first eight rows are the corner experiments of the basis cube, the next six rows the so-called axial points, and the last three rows the replicated center-points. Noteworthy is the fact that this design is slightly distorted. Because of minor experimental obstacles, it was difficult to adjust the three factors so that they exactly matched the specifications of the CCF design. For instance, the levels 240 and 284 were set for Air in the factor specification, but in reality small deviations around these targets were observed. Such small deviations from the factor specification are found also for the second factor, EGR%, and for the third factor, NeedleLift. However, for NeedleLift the deviation is found only at the center-level. Interestingly, we can use the condition number for evaluating the magnitude of the design distortion. The condition number for the conventional CCF in three factors is 4.438, but for the distorted truck engine design it amounts to 4.508, that is, an increase of 0.07. Changes in the condition number smaller than 0.5 are regarded as insignificant. We can therefore infer that the design distortion is almost non-existent, and will not compromise the validity or usefulness of the work.

Introduction to response surface methodology (RSM) designs

We have now completed the problem formulation phase of the truck engine application, and it is appropriate to take a closer look at the properties of the CCF design we have employed. The CCF design is often classified as an RSM design. Traditionally, RSM has been the acronym for *response surface methodology*, reflecting the predominant view that extensive use of response surface plots is advantageous for finding an optimal point. In more recent years, however, the re-interpretation *response surface modelling* has become more prevalent. This name emphasizes the significance of the modelling and understanding part, that is, the importance of verifying the reliability of a regression model prior to its conversion to response contour plots or response surface plots.

In RSM it is important to get good regression models and in general higher demands are placed on these models than on screening models. Consequently, there are certain fundamental demands which must be met by an RSM design. Good RSM designs have to allow the estimation of the parameters of the model with low uncertainty, which means that we want the confidence intervals of the regression coefficients to be as narrow as possible. Good RSM designs should also give rise to a model with small prediction error, and permit a judgement of the adequacy of this model. This latter aspect means that the design must contain replicated experiments enabling the performance of a lack of fit test. In addition, good RSM designs should encode as few experiments as possible. There are several classical RSM design families which meet these demands, and we will outline three of these. These are the central composite, Box-Behnken, and three-level full factorial design families. Computer generated D-optimal designs for irregular experimental regions may be tailored to fulfill such requirements, as well.

The CCC design in two factors

In the current section our intention is to describe the family of composite designs. Box-Behnken, three-level full factorial and D-optimal designs are discussed in Chapter 16. The composite designs are natural extensions of the two-level full and fractional factorial designs. We commence with the central composite circumscribed, CCC, design in two factors. Figure 15.15 displays this type of design. The CCC design consists of three building blocks, (i) regularly arranged corner (factorial) experiments of a two-level factorial design,

(ii) symmetrically arrayed star points located on the factor axes, and (iii) replicated center-points. Hence, it is a *composite* design.
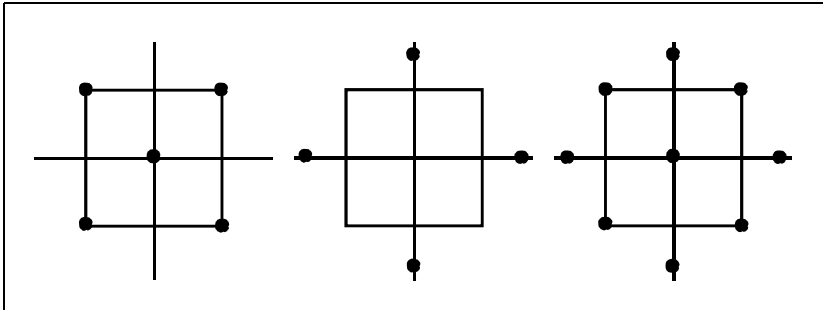
Figure 15.15: The CCC design in two factors.

Figure 15.16 gives the worksheet of an 11 run CCC design in two factors. Here, the first four rows represent the corner experiments, the next four rows the star points, and the last three rows the replicated center-points. The star points, also denoted axial points, represent the main difference between factorial and composite designs. Thanks to them, all factors are investigated at five levels with the CCC design. This makes it possible to estimate quadratic terms with great rigor. We can see in Figures 15.15 and 15.16 that the corner experiments and the axial experiments are all situated on the circumference of a circle with radius 1.41, and therefore the experimental region is symmetrical.

	Exp No	Exp Name	Run Order	Incl/Excl	x1	x2
1	1	N1	6	Incl ▼	−1	−1
2	2	N2	11	Incl ▼	1	−1
3	3	N3	9	Incl ▼	−1	1
4	4	N4	10	Incl ▼	1	1
5	5	N5	3	Incl ▼	−1.414	0
6	6	N6	2	Incl ▼	1.414	0
7	7	N7	7	Incl ▼	0	−1.414
8	8	N8	4	Incl ▼	0	1.414
9	9	N9	8	Incl ▼	0	0
10	10	N10	1	Incl ▼	0	0
11	11	N11	5	Incl ▼	0	0

Figure 15.16: Worksheet of CCC design in two factors.

With two, three and four factors, the factorial part of the CCC design corresponds to a full two-level factorial design. With five or more factors a fractional factorial design of resolution V is utilized. This means that with any CCC design, all linear, two-factor interaction, and quadratic model terms are estimable.

The CCC design in three factors

The CCC design in three factors is constructed in a fashion similar to that of the two factor analogue. Figure 15.17 displays its three components, which are (i) eight corner experiments, (ii) six axial experiments, (iii) and, at least, three replicated center-points.
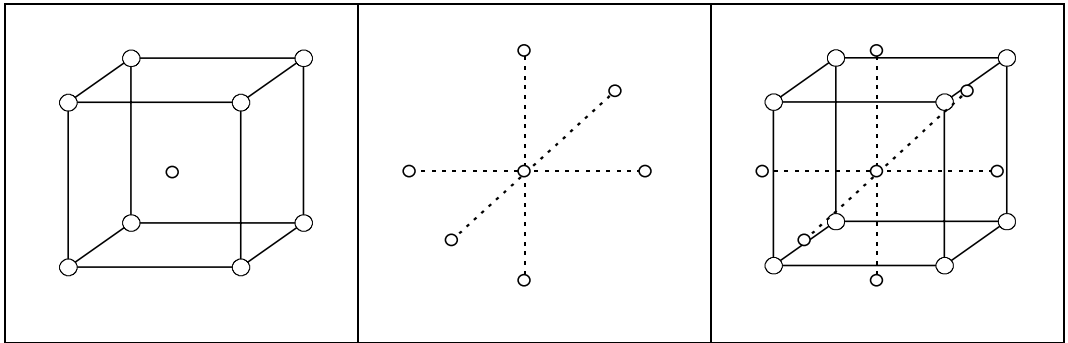
Figure 15.17: The CCC design in three factors.

The worksheet of such a 17 run design is displayed in Figure 15.18. Evidently, with three factors, the corner and axial experiments approximate the surface of a sphere and hence the experimental region is symmetrical. In a similar manner, four-factor and five-factor CCC designs define hyper-spherical experimental arrangements.

	Exp No	Exp Name	Run Order	Incl/Excl	x1	x2	x3
1	1	N1	2	Incl	-1	-1	-1
2	2	N2	7	Incl	1	-1	-1
3	3	N3	13	Incl	-1	1	-1
4	4	N4	17	Incl	1	1	-1
5	5	N5	16	Incl	-1	-1	1
6	6	N6	14	Incl	1	-1	1
7	7	N7	10	Incl	-1	1	1
8	8	N8	9	Incl	1	1	1
9	9	N9	5	Incl	-1.682	0	0
10	10	N10	1	Incl	1.682	0	0
11	11	N11	4	Incl	0	-1.682	0
12	12	N12	11	Incl	0	1.682	0
13	13	N13	8	Incl	0	0	-1.682
14	14	N14	3	Incl	0	0	1.682
15	15	N15	6	Incl	0	0	0
16	16	N16	12	Incl	0	0	0
17	17	N17	15	Incl	0	0	0

Figure 15.18: Worksheet of CCC design in three factors.

The CCF design in three factors

The CCC design prescribes experiments, the axial points, whose factor values are located outside the low and high settings of the factor definition. Sometimes it is not possible to carry out this kind of testing. When it is desirable to maintain the low and high factor levels, and still perform an RSM design, the central composite face-centered (CCF) design is a viable alternative. This design in three factors is shown in Figure 15.19 and for comparison its CCC counterpart is plotted in Figure 15.20.
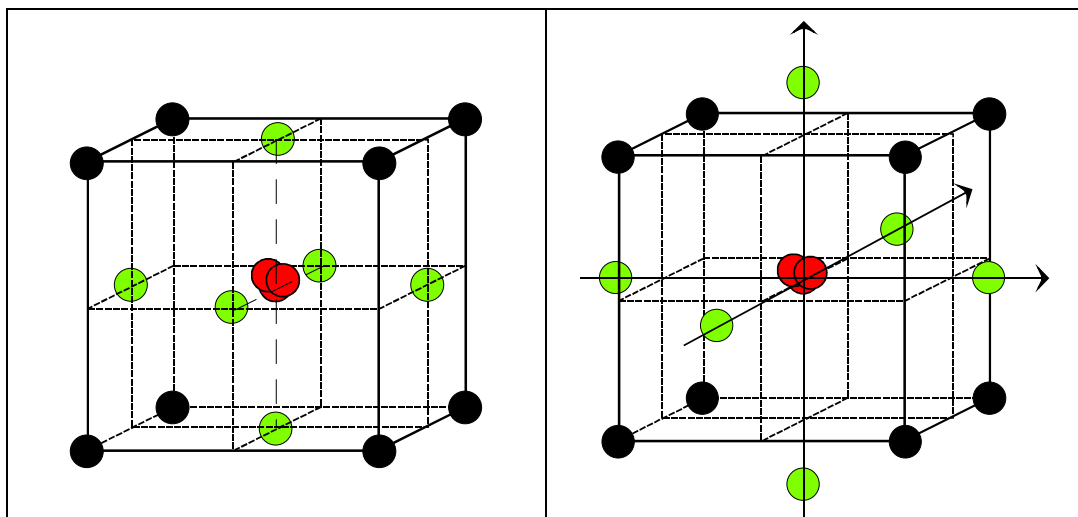
Figure 15.19: *(left) CCF design in three factors.*
Figure 15.20: *(right) CCC design in three factors.*

In the CCF design, the axial points are centered on the faces of the cube. This implies that all factors have three levels, rather than five, and that the experimental region is a cube, and not a sphere. CCF is the recommended design choice for pilot plant and full scale investigations. Although it corresponds to only three levels of each factor, it still supports a quadratic model, because it contains an abundance of experiments. Theoretically, the CCF design is slightly inferior to the CCC design. Given the same settings of low and high levels of the factors, the CCC design spans a larger volume than does the CCF design. Five levels of each factor also means that the CCC design is better able to capture strong curvature; even a cubic response behavior may be modelled. Figures 15.21 and 15.22, which provide the correlation matrices of the CCF and CCC designs, illustrate another argument in favor of the CCC design. It is seen that the quadratic model terms are less correlated in the CCC than CCF case, which is beneficial when it comes to regression modelling.

Figure 15.21: *(left) Correlation matrix of the CCF design in three factors.*
Figure 15.22: *(right) Correlation matrix of the CCC design in three factors.*

In summary, the CCF design is very good, and some of its properties are very tractable in pilot plant and large-scale operations. Interestingly, since the CCF and CCC designs require the same number of experiments, it is possible to tailor-make a design by adjusting the distance of the axial points from the design center. Each such CCF/CCC-hybrid is a very good RSM design, but one must remember to check the correlation matrix.

Overview of composite designs

Usually, RSM designs, like CCC and CCF, are used in conjunction with two to five factors. Figure 15.23 provides a tabulated overview of how many experiments are necessary in these cases. For comparative purposes, the table is also extended to cover the situation with six and seven factors. We see that it is possible to explore as many as five factors in as few as 29 experiments. These 29 experiments consist of 16 corner experiments, 10 axial experiments, and three replicated center-points. However, when moving up to six factors, there is a huge increase in the number of experiments, which effectively means that RSM should be avoided in this case.

Number of factors	Number of experiments
2	8 + 3
3	14 + 3
4	24 + 3
5	26 + 3
6	44 + 3
7	78 + 3

Figure 15.23: Overview of composite designs in 2 to 7 factors.

Observe that the CCC and CCF designs in two, three, and four factors are based on two-level full factorial designs, whereas in the case of five or more factors two-level fractional factorial designs are employed as the design foundation. The recommended number of replicated center-points has been set to three for all designs. This is the lowest number needed to provide a reasonable estimate of the replicate error. For various reasons, however, it is sometimes desirable to increase the number of center-points. By adjusting the number of center-points it is possible to manipulate certain properties of the design, such as, its orthogonality and the overall precision of its predictions. This is an advanced topic that will not be pursued in this course. However, the important thing to remember is that for all practical purposes, three replicated center-points is an appropriate choice. With this discussion in mind, we will, after a quiz break, continue with the truck engine application and the analysis of this data-set.

Quiz I

Please answer the following questions:

Which questions are typically asked in optimization?

Which kind of regression model is usually fitted in optimization?

What is RSM?

What is a composite design?

What are the differences between the CCC and CCF designs?

Which design is recommended in full scale applications?

How many factors should, at maximum, be explored with an RSM design?

How many experiments are needed to perform an RSM study of two factors? Three, four and five factors?

Truck engine application

Evaluation of raw data

We start the evaluation of the raw data by inspecting the replicate error of each response. Figures 15.24 – 15.26 display plots of replications for the three responses Fuel, NOx, and Soot.
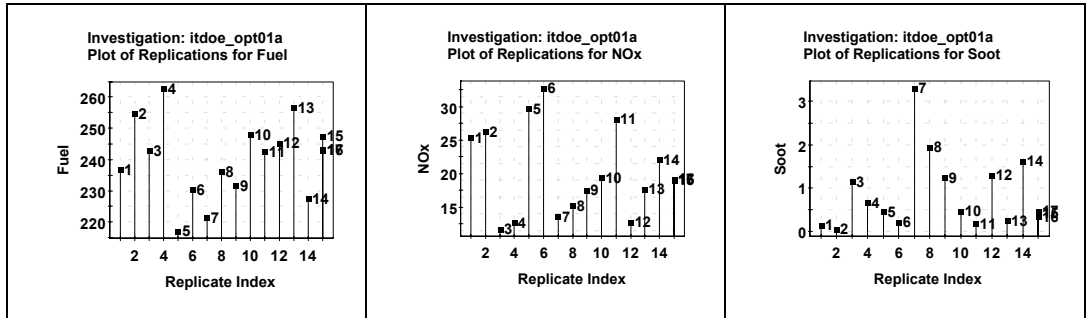
Figure 15.24: *(left) Replicate plot of Fuel.*
Figure 15.25: *(middle) Replicate plot of NOx.*
Figure 15.26: *(right) Replicate plot of Soot.*

We can see that the replicate errors are small in all cases, which is good for the forthcoming regression analysis. However, we can also see that for Soot we have one measurement, #7, which is much larger than the others. This means that we must expect the distribution of Soot to be positively skewed. That this is indeed the case is seen in the next triplet of figures, 15.27 – 15.29, which provide the histogram of each response.
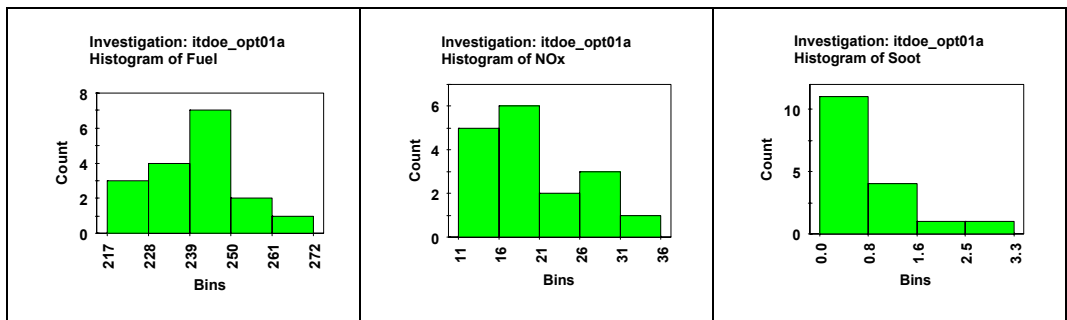
Figure 15.27: *(left) Histogram of Fuel.*
Figure 15.28: *(middle) Histogram of NOx.*
Figure 15.29: *(right) Histogram of Soot.*

Figure 15.29 reveals the need for log-transforming Soot. The appearance of the histogram and the replicate plot after log-transformation of Soot is given by Figures 15.30 and 15.31. It is clear that the log-transformation is warranted. A final check of the condition number in Figure 15.32 indicates that we have a pertinent design. Recall that classical RSM designs in 2-5 factors have condition number in the range 3-8.
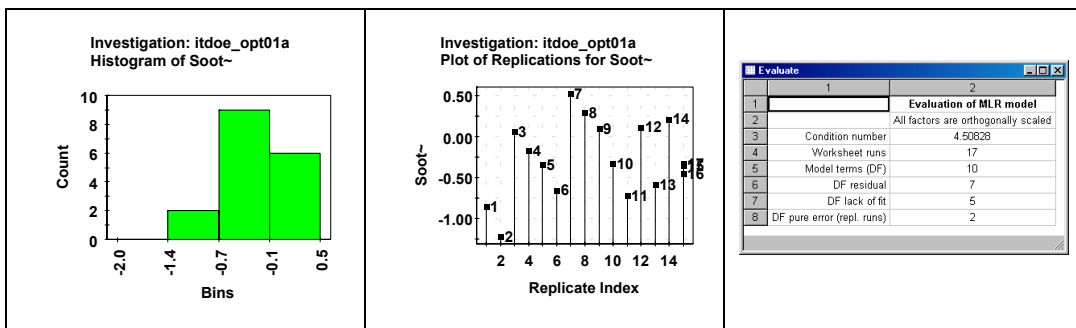
Figure 15.30: *(left) Histogram of Soot – after log-transformation.*
Figure 15.31: *(middle) Replicate plot of Soot – after log-transformation.*
Figure 15.32: *(right) Condition number evaluation.*

Regression analysis

We will now fit a quadratic model with 10 terms to each response. Each model will have one constant, three linear, three quadratic, and three two-factor interaction terms. The overall result of the model fitting is displayed in Figure 15.33. It is seen that the predictive power ranges from good to excellent. The Q^2 values are 0.93, 0.97, and 0.75, for Fuel, NOx, and Soot, respectively.

Fuel	DF	SS	MS (variance)	F	p	SD
Total	17	985053	57944.293			
Constant	1	982647.5	982647.5			
Total Corrected	16	2405.5	150.344			12.261
Regression	9	2373.759	263.751	58.166	9.48E-06	16.24
Residual	7	31.741	4.534			2.129
Lack of Fit (Model Error)	5	19.501	3.9	0.637	0.704	1.975
Pure Error (Replicate Error)	2	12.24	6.12			2.474

N = 17 Q2 = 0.9283 CondNo = 4.5083
DF = 7 R2 = 0.9868 Y-miss = 0
R2Adj = 0.9698 RSD = 2.1294
ConfLev= 0.95

NOx	DF	SS	MS (variance)	F	p	SD
Total	17	7522.213	442.483			
Constant	1	6876.699	6876.699			
Total Corrected	16	645.514	40.345			6.352
Regression	9	644.264	71.585	400.713	1.19E-08	8.461
Residual	7	1.251	0.179			0.423
Lack of Fit (Model Error)	5	1.147	0.229	4.421	0.195	0.479
Pure Error (Replicate Error)	2	0.104	0.052			0.228

N = 17 Q2 = 0.9731 CondNo = 4.5083
DF = 7 R2 = 0.9981 Y-miss = 0
R2Adj = 0.9956 RSD = 0.4227

Log(Soot)	DF	SS	MS (variance)	F	p	SD
Total	17	4.637	0.273			
Constant	1	1.328	1.328			
Total Corrected	16	3.309	0.207			0.455
Regression	9	3.166	0.352	17.217	5.55E-04	0.593
Residual	7	0.143	0.02			0.143
Lack of Fit (Model Error)	5	0.137	0.027	8.494	0.109	0.165
Pure Error (Replicate Error)	2	0.006	3.22E-03			0.057

N = 17 Q2 = 0.7486 CondNo = 4.5083
DF = 7 R2 = 0.9568 Y-miss = 0
R2Adj = 0.9012 RSD = 0.1429

Figure 15.33: *(upper left) Summary of fit of initial modelling.*
Figure 15.34: *(upper right) ANOVA of Fuel.*
Figure 15.35: *(lower left) ANOVA of NOx.*
Figure 15.36: *(lower right) ANOVA of log Soot.*

Since this is RSM and the final goal is optimization, we would prefer a higher Q^2 of Soot. In order to see what may be done to achieve this, we proceed with the other diagnostic tools. The results of the second diagnostic tool, the analysis of variance, are summarized in Figures 15.34 – 15.36. Remembering that the upper p-value should be smaller than 0.05 and the lower p-value larger than 0.05, we realize that none of the models exhibits any significant lack of fit. This is good. The outcome of the third diagnostic tool, the normal probability plot of residuals, is shown in Figures 15.37 – 15.39.
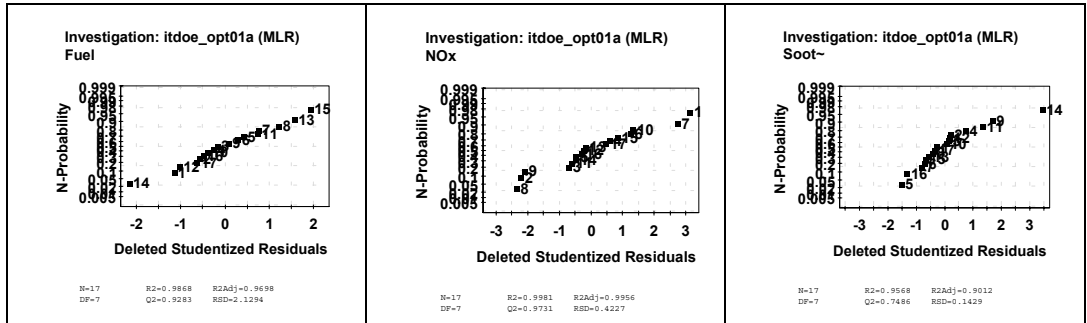
Figure 15.37: *(left) N-plot of Fuel residuals.*
Figure 15.38: *(middle) N-plot of NOx residuals.*
Figure 15.39: *(right) N-plot of Soot residuals.*

Primarily, the three N-plots of residuals suggest that attention should be given to experiment #14, as it deviates a little with respect to Soot. This deviation might be the cause of the comparatively low Q^2 for Soot. However, it is important to underscore that experiment 14 is *not* a strong outlier. This is inferred from the high R^2 value of the response. Another possible reason as to why Q^2 is low for Soot might be that the regression model contains irrelevant terms. This may be checked through a bar chart of its regression coefficients. Figures 15.40 – 15.42 provide plots of the regression coefficients for each model. After some scrutiny of these plots, we can detect two model terms which are not significant for any one of the three responses. These are the Air*EGR and Air*NL two-factor interactions. Thus, in the model refinement step, it is logical to remove these and explore the modelling consequences.
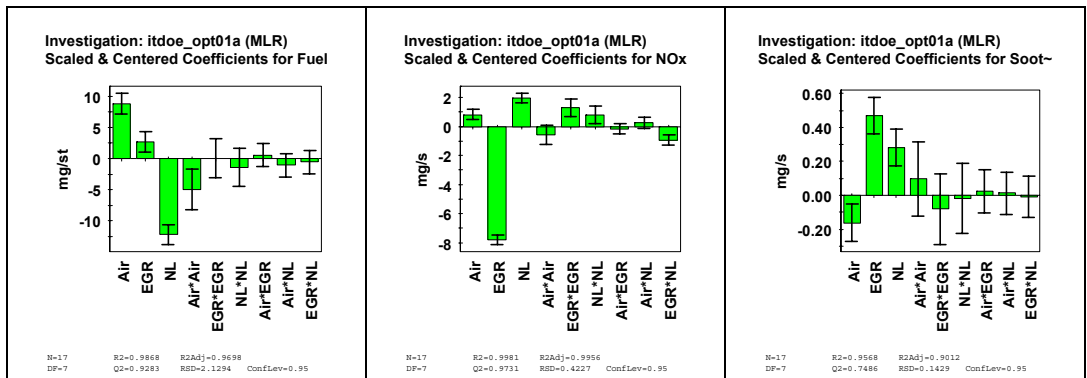
Figure 15.40: *(left) Regression coefficients of Fuel model.*
Figure 15.41: *(middle) Regression coefficients of NOx model.*
Figure 15.42: *(right) Regression coefficients of Soot model.*

Model refinement

Upon deletion of the two two-factor interactions and refitting of the models, we obtained even better results. The summary of fit plot shown in Figure 15.43 shows that all three Q^2 values have increased, and now amount to 0.94, 0.99, and 0.84 for Fuel, NOx, and Soot, respectively. Thus the model refinement seems rewarding. However, before we can be sure that this model pruning is justifiable, we must consider the remaining diagnostic tools.

Fuel	DF	SS	MS (variance)	F	p	SD
Total	17	985053	57944.293			
Constant	1	982647.5	982647.5			
Total Corrected	16	2405.5	150.344			12.26
Regression	7	2363.388	337.627	72.156	3.47E-07	18.37
Residual	9	42.112	4.679			2.163
Lack of Fit (Model Error)	7	29.88	4.269	0.698	0.699	2.066
Pure Error (Replicate Error)	2	12.232	6.116			2.473

N = 17	Q2 = 0.9372	CondNo = 4.498
DF = 9	R2 = 0.9825	Y-miss = 0
	R2Adj = 0.9689	RSD = 2.1631
		ConfLev= 0.95

NOx	DF	SS	MS (variance)	F	p	SD
Total	17	7522.213	442.483			
Constant	1	6876.699	6876.699			
Total Corrected	16	645.514	40.345			6.352
Regression	7	643.59	91.941	429.941	1.24E-10	9.589
Residual	9	1.925	0.214			0.462
Lack of Fit (Model Error)	7	1.821	0.26	5.002	0.177	0.51
Pure Error (Replicate Error)	2	0.104	0.052			0.228

N = 17	Q2 = 0.9868	CondNo = 4.498
DF = 9	R2 = 0.997	Y-miss = 0
	R2Adj = 0.9947	RSD = 0.4624

Log(Soot)	DF	SS	MS (variance)	F	p	SD
Total	17	4.637	0.273			
Constant	1	1.328	1.328			
Total Corrected	16	3.309	0.207			0.455
Regression	7	3.16	0.451	27.376	2.26E-05	0.672
Residual	9	0.148	0.016			0.128
Lack of Fit (Model Error)	7	0.142	0.02	6.313	0.144	0.142
Pure Error (Replicate Error)	2	0.006	3.21E-03			0.057

N = 17	Q2 = 0.8376	CondNo = 4.498
DF = 9	R2 = 0.9551	Y-miss = 0
	R2Adj = 0.9203	RSD = 0.1284

Figure 15.43: *(upper left) Summary of fit of refined modelling.*
Figure 15.44: *(upper right) ANOVA of Fuel – after model refinement.*
Figure 15.45: *(lower left) ANOVA of NOx – after model refinement.*
Figure 15.46: *(lower right) ANOVA of log Soot – after model refinement.*

The relevant ANOVA tables and N-plots of residuals are shown in Figures 15.44 – 15.49. Everything looks good in these plots, except for the deviating behavior of experiment 14 in Figure 15.49. However, because of the very good R^2 and Q^2 of this model, we can conclude that #14 is a weak outlier, which does not influence the model decisively. We may regard these models as the best we can accomplish with the current data.
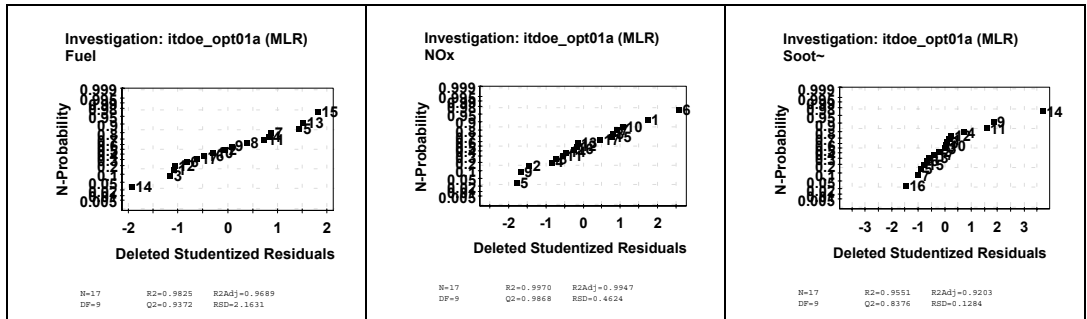
Figure 15.47: *(left) N-plot of Fuel residuals – after model refinement.*
Figure 15.48: *(middle) N-plot of NOx residuals – after model refinement.*
Figure 15.49: *(right) N-plot of Soot residuals – after model refinement.*

The models may be interpreted with the help of the regression coefficient plots shown in
Figures 15.50 – 15.52. Because the coefficient patterns are quite variable among the models,
the interpretation in this case is not straightforward. The overall picture is that linear terms
dominate over terms of higher order. The main effects of Air and NeedleLift are influential
for regulating Fuel, the main effect of EGR most meaningful for NOx, and the main effects
of EGR and NeedleLift most important for Soot.
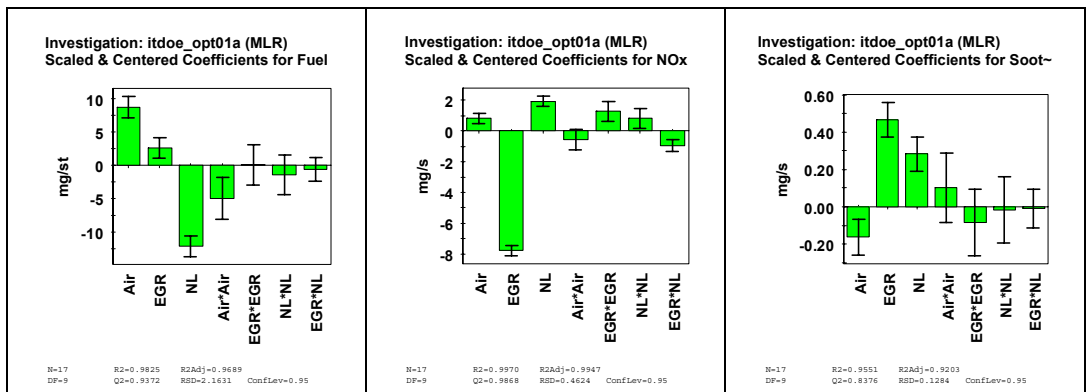
Figure 15.50: *(left) Regression coefficients of Fuel model.*
Figure 15.51: *(middle) Regression coefficients of NOx model.*
Figure 15.52: *(right) Regression coefficients of Soot model.*

Use of model

When trying to extract concrete information from our models for deciding what to do next,
it is convenient to use the overview plot of the regression coefficients plotted in Figure
15.53.
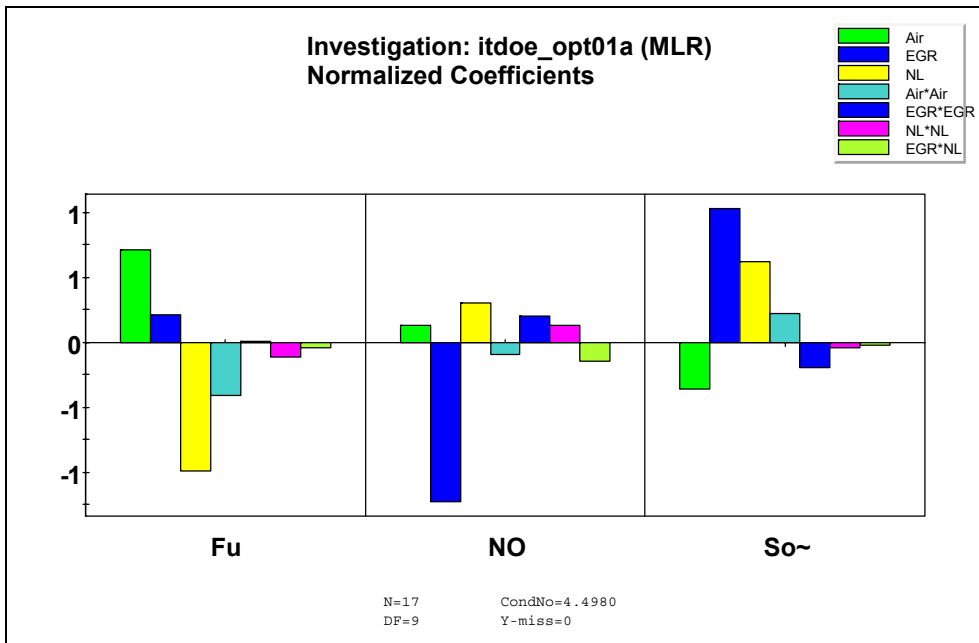
Figure 15.53: Coefficient overview plot of final truck engine models.

It is not immediately clear from Figure 15.53 which two factors should be put on the axes in the contour plots, and which factor should be held constant. One way to attack this problem could be as follows: We can see that Air, in principle, is only influential for Fuel. Since we want to get fuel consumption as low as possible, we should perhaps fix Air at its low level, and then let the other two factors vary. Such response contour plots are plotted in Figure 15.54. The solid arrows indicate where the desirabilities of the three responses are met.
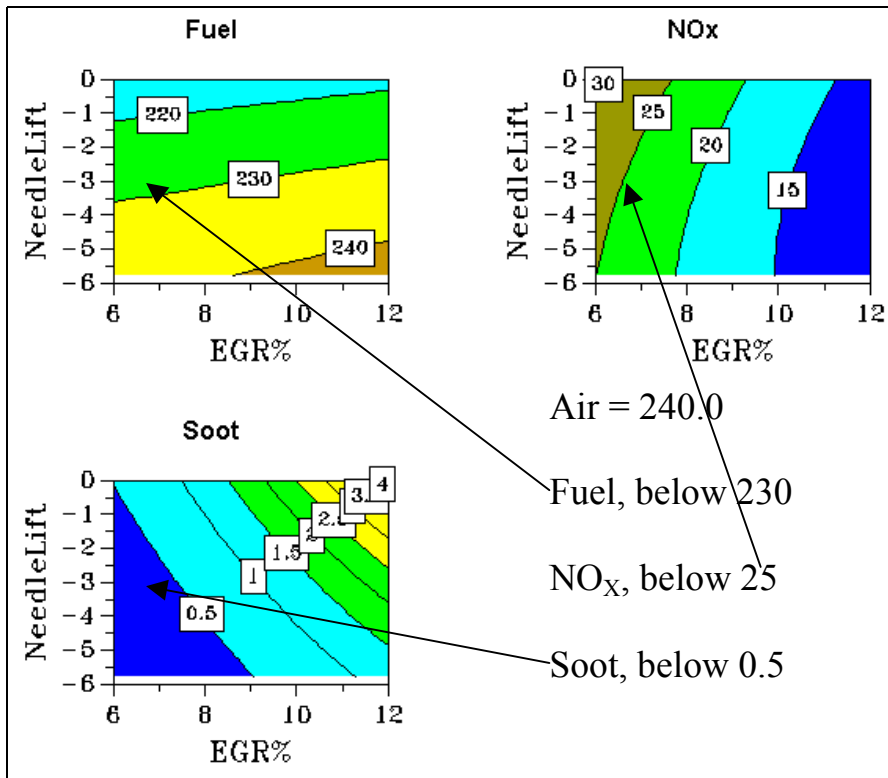
Figure 15.54: *Response contour triplets of final truck engine model.*

Obviously, we should strive to stay somewhere in the lower left-hand corner, that is, use low levels of all three factors. The strong message conveyed by these contour plots is that the specified response profile is reasonable and will realistically be met within the mapped region. In principle, it is possible to carry on this procedure by using other triplets of response contour plots, in order to find the optimal point. However, this is both tedious and time-consuming. Fortunately, it is possible to automate this procedure with the optimizer in MODDE. We will investigate this option in a moment, but first provide some more theory relating to what to do after response surface modelling. Subsequently, we shall return to the analysis of the truck engine data.

What to do after RSM

Introduction

What to do after RSM depends on the results obtained. There are three ways in which we can proceed. The first way is appropriate when one of the experiments fulfils the goals stated in the problem formulation. This corresponds to the ideal case, and in principle only a couple of immediate verification experiments are needed to establish the usefulness of this factor combination. A second possibility involves regression modelling and use of the model for predicting the location of a new promising experimental point. This is the most

common of the three options, and is deployed when none of the performed design runs are really close to fulfilling the experimental goals.

The third, and least frequently used, option is taken when a few extra experiments are needed to supplement the existing RSM design. For instance, it may be the case that a rather good quadratic model has been acquired. But, during the course of the data analysis, certain weaknesses of the quadratic model have been exposed, and it appears that the only remedy is the fitting of a cubic model term in one of the dominant factors. In this instance, it is possible to add to the parent design a couple of supplementary design runs, laid out D-optimally, which would enable the estimation of the lurking cubic term. This was a short overview of the three primary measures that might be undertaken after RSM. In the following, we shall focus solely on the second situation, as it represents the most common action undertaken after RSM.

Interpolation & Extrapolation

When entering the stage of optimization, the user has a great deal of insight into the investigated system compared with the screening situation. Now, the investigator *knows* that the optimum must be close to the investigated region. The optimum is not far away, as may be the case in screening. This relative proximity to the optimal point implies that it is possible to use the RSM model for predicting the probable location of such an optimum. It is not – as in screening – necessary to employ any gradient technique for finding the optimal experimental region.

When we use the model for making predictions of the responses, and the predictions are made inside the explored experimental domain, we say that we conduct *interpolation*. Analogously, when we predict outside the region, we say that we perform *extrapolation*. Regardless of whether interpolation or extrapolation is taking place, it is important that the RSM model is predictively relevant. Earlier we presented a number of raw data evaluation tools and regression modelling diagnostic tools, which are useful for verifying that the model is reliable. When our model has been found relevant, the question is *in which direction of the experimental region do we make the predictions*? Without any rational strategy for this we could go on for a long time, because there are infinitely many factor combinations at which we could calculate response values. Fortunately, MODDE contains an optimization routine, which will automatically browse through a large number of factor combinations, and in the end select only a few at which fruitful predictions have been made.

Automatic search for an optimal point

The MODDE optimizer was introduced in Chapter 13 and we will now give a short resume of what was written there. Firstly, the roles of the factors must be given. A factor can be allowed to change its value freely in the optimization, or it may be fixed at a constant value. In addition, for the variable factors, low and high variation limits must be set. Secondly, certain criteria concerning the responses must be given. A response variable may either be maximized, minimized, directed towards an interval, or excluded from the optimization.

Subsequent to these definitions, the optimizer will use the fitted regression model and the experiments already performed to compute eight starting points for simplexes. In principle, the optimizer then uses these simplexes, together with the fitted model, to optimize the desirabilities, or goals, of the individual responses. These goals are taken from the criteria specified for the responses. Figure 15.55 shows one simplex in action. For simplicity this simplex is laid out in a two-dimensional factor space.
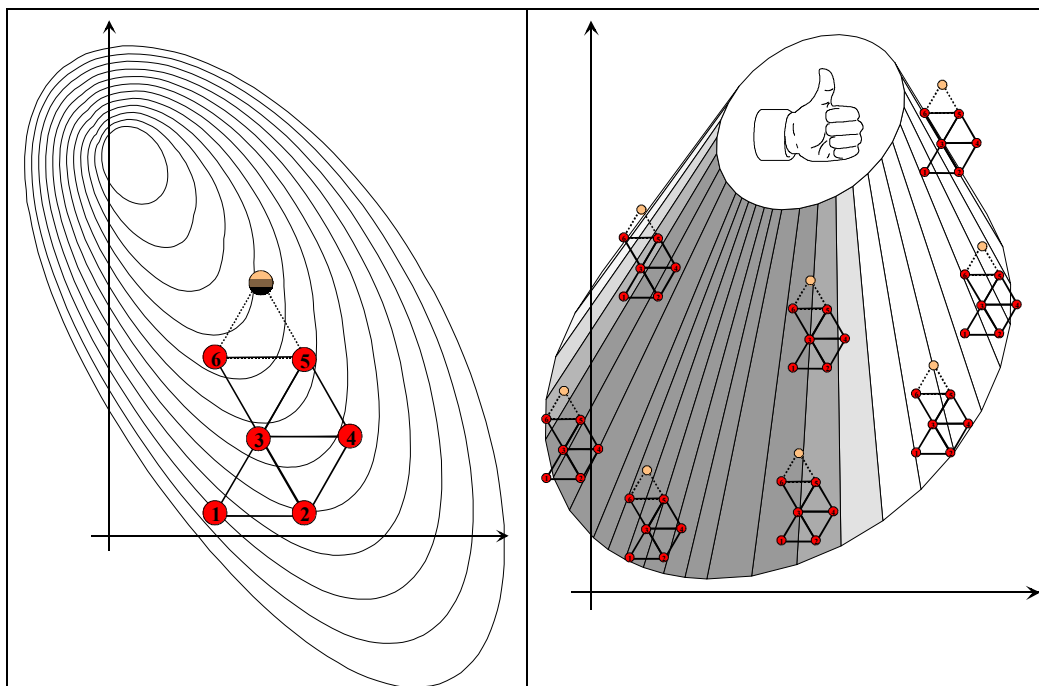
Figure 15.55: *(left) An illustration of the simplex methodology.*
Figure 15.56: *(right) Different simplexes get different starting points in factor space.*

The first simplex consists of the three experiments numbered 1, 2, and 3. Now, the idea is to mirror the worst experiment, here #1, through the line connecting the two best experiments, here # 2 and 3, and do a new experiment, here #4, at the position where the mirrored experiment "hit" the response contour plot. This means that the new simplex will consist of runs 2, 3, and 4. In the next cycle, the worst experiment is mirrored in the line connecting the two best experiments, and so on. This is repeated until the peak of the elliptical mountain is reached by one simplex. We emphasize that this is only a simplified picture of how the simplex methodology works. In reality, when a new simplex is formed through a reflection, it may be reduced or enlarged in size in comparison with the precursor simplex.

The optimizer will simultaneously start up to eight simplexes, from different locations in the factor space. Eight simplexes are initiated in order to avoid being trapped at a local minimum or maximum. This is illustrated in Figure 15.56. The co-ordinates for the simplex starting points are taken from the factors with the largest coefficients in the regression model. At convergence, each simplex will display the best factor settings, and it is then possible to compute response values at each point. Observe that the drawing in Figure 15.56 is a schematic valid when striving for higher response values. In reality, one could also look for a minimum, and the relevant response function would then correspond to a hole or a valley.

MODDE optimizer applied to the truck engine data

We will now describe what happened when applying the optimizer to the truck engine application. It is a good practice to conduct the optimization in stages. We recommend that

the first phase is conducted according to the principles of interpolation, even though extrapolation might sometimes (in screening) be the final result. In interpolation, the low and high limits of the factors are taken from the factor definition. This is shown in Figure 15.57.

	Factor	Role		Value	Low Limit	High Limit
1	Air	Free	▼		240	284
2	EGR%	Free	▼		6	12
3	NeedleLift	Free	▼		-5.78	0

Figure 15.57: *Factor settings in optimizer study of truck engine application.*

Next, it is necessary to give the desirability, or goal, of each response, that is, to specify a desired response profile. We see the criterion and goal set for each response in Figure 15.58. This table is understood as follows. All three responses, Fuel, NOx, and Soot are to be minimized. We want Fuel to be lower than 230, NOx lower than 25, and Soot below 0.5. Technically, these minimization requirements are specified by assigning two reference values for each response. One value is the target value, which represents the numerical value of each response towards which the optimization is driven. The other value is the maximum value which we can accept should the corresponding target value be unattainable. Furthermore, it is also possible to assign weighting coefficients to each response. Such weights may range from 0.1 to 1.0. Here, we assign the same priority to all three responses, by stating that each weight should be 1.0. This weighting procedure is an advanced topic that is not treated further in this course.

	Response	Criteria		Weight	Min	Target	Max
1	Fuel	Minimize	▼	1		225	230
2	NOx	Minimize	▼	1		10	25
3	Soot	Minimize	▼	1		0.05	0.5

Figure 15.58: *Response desirabilities in optimizer study of truck engine application.*

First optimization - Interpolation

Now we are ready to begin the optimization. In doing so, the optimizer will first compute appropriate starting co-ordinates, factor combinations, for launching the eight simplexes. Five starting points are derived by laying out a 2^{3-1} fractional factorial design with one center-point in the three model factors. This design corresponds to the five first rows in the spreadsheet of Figure 15.59. In addition, the optimizer browses through the existing worksheet and identifies those three experiments which are closest to the specified response profile. These are the three runs listed in rows 6, 7 and 8.

Figure 15.59: Starting co-ordinates of the simplexes in the first optimization round.

On pressing the single-headed arrow, the eight simplexes will start their search for a factor combination, at which the predicted response values are as similar to the desired response profile as possible. The outcome of the eight simplexes are seen in Figure 15.60. All proposed factor combinations are different, but some are rather similar. The success of each simplex may be deduced from the numerical values of the rightmost column. This logD value represents a weighted average of the individual response desirabilities. It may be used to evaluate how the optimization proceeds. A positive logD is bad. When logD is zero, all responses are predicted to be between their assigned target and maximum value. This is a good situation. Better still is when logD is negative, and –10 is the lower, optimal, limit. Hence, in our case we are doing rather well, and it seems possible to find a factor combination where all three responses will be below the critical limit. Therefore, we are close to detecting an optimal point *within* the investigated experimental region. This means that extrapolation is unnecessary.

Figures 15.60: Optimizer results of the first optimization round.

Second optimization – New starting points around first selected run

One problem which might compromise the validity of the simplex optimization approach is the risk of being trapped by a local minimum or maximum, and thereby missing the global optimal point. This may be circumvented by taking the simplex predicted to be the best, and from its factor combination generate new starting points for simplexes. We decided to exploit this option.

Figure 15.61: Starting co-ordinates of the simplexes in the second optimization round.

The initially predicted best point is shown as the fifth row in the spreadsheet of Figure 15.61. Around this point, four new starting points, rows 1-4, were defined. These co-ordinates were identified by moving away from the mother simplex, row 5, to a distance corresponding to 20% of the factor ranges. By executing the four new simplexes and the old one, the results displayed in Figure 15.62 were obtained.

Figure 15.62: Optimizer results of the second optimization round.

Evidently, all five points are predicted to meet our experimental goals. In fact, by neglecting some small fluctuations among the decimal digits, we find that the five simplexes have converged to approximately the same point, that is, Air = 240, EGR% $\approx$ 7.0, and NeedleLift $\approx$ -3.20. First of all, this means that these simplexes were not trapped by a local response phenomenon. Secondly, it indicates that the response surfaces do not display any dramatic changes in the vicinity of the identified optimal point.

Graphical evaluation of optimization results

It is possible to investigate graphically the situation around the predicted optimal point, just to see if the results are sensitive to small changes in the factor values. Figure 15.63 provides

the response contour plots which were obtained by centering around EGR% = 7.0 and NeedleLift = -3.2, and holding Air = 240 fixed.
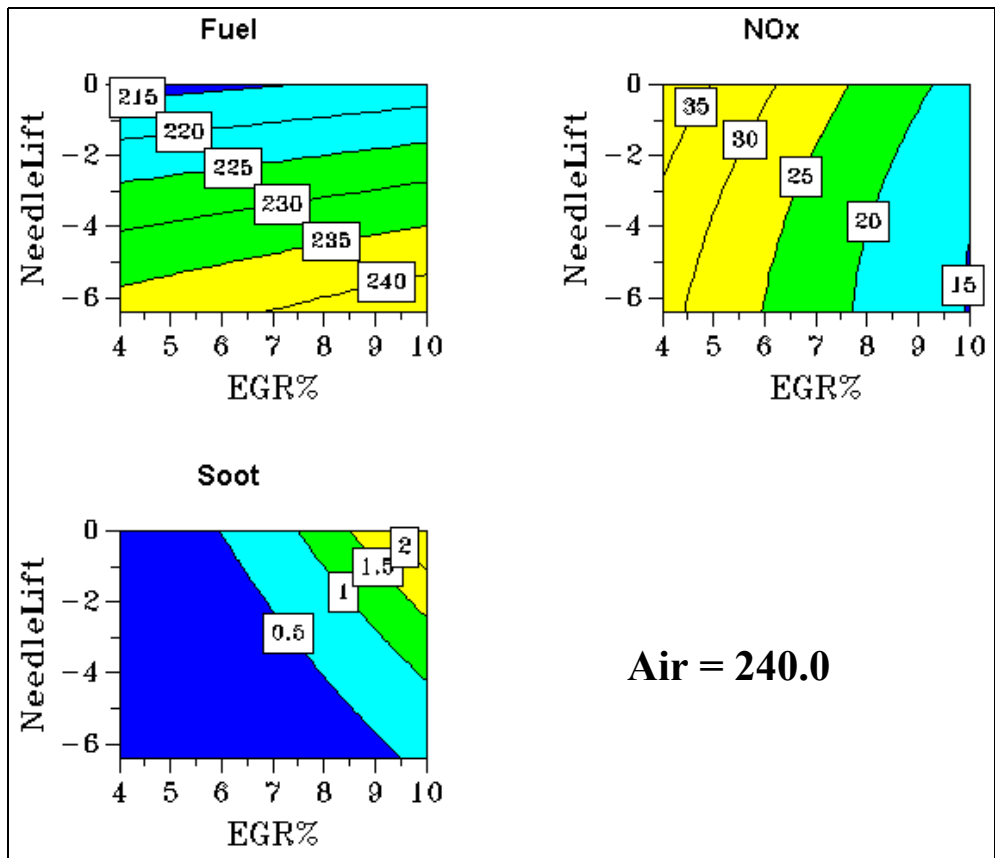
Figure 15.63: Graphical exploration of the situation around the predicted optimal point.

Evidently, there are no dramatic changes occurring in the predicted response functions. Rather, there are slow and controlled response changes. Hence, it appears relevant to conduct a few verifying experiments with the setting Air = 240, EGR% = 7.0, and NeedleLift = -3.2. The predicted response values at this point are shown in Figure 15.64 together with the 95% confidence intervals. It is seen that each point estimate fulfils the experimental goals. The investigator Sven Ahlinder carried out verifying experiments and found that the proposed optimum was well-founded.

Figure 15.64: Confidence intervals (95%) of predicted response values at the optimal point.

Summary of the truck engine application

The application of DOE in the construction of a well-behaved truck engine was successful. With this approach it was possible not only to understand the mechanisms involved in the combustion, but also to identify an optimal factor combination at which low fuel consumption and minimal pollution prevailed. From an educational point of view, the truck engine application is good for several reasons. For instance, it clearly shows that it is possible to lay out an optimization design and find an optimal point within the investigated experimental region. In this case, this was possible thanks to carefully conducted screening work.

Furthermore, this data set also allowed us to study how tools for evaluation of raw data, tools for regression analysis, and tools for model use, were put into practice in a real situation. It should be clear from this application that the process of deriving the best possible regression model is often iterative in nature, and many repetitive modelling cycles are often needed before a final model can be established. This modelling step is of crucial importance for the future relevance of a DOE project. If an investigator fails to ensure the quality of a regression model, this may have negative consequences at later stages when the model is used for predictions. Hence, strong emphasis is placed on finding the most reliable and useful model.

We also used the truck engine study for reviewing what options are available after RSM. The most common action after RSM is to use the developed model for optimization. With the software optimizer we rapidly identified an optimal point with the factor combination Air = 240, EGR% = 7.0, and NeedleLift = -3.2, at which the fuel consumption was predicted to be below 230, and the levels of NOx and Soot in the car exhaust gases modelled to be below 25 and 0.5, respectively. The principal experimenter Sven Ahlinder later verified this point experimentally.

Summary of composite designs

The composite designs CCC and CCF are natural extensions of the two-level full and fractional factorial designs, which explains their popularity in optimization studies. In fact, it is possible, by rather simple measures, to extend a screening design to become an RSM design. A composite design consists of three building blocks, (i) regularly arranged corner experiments of a two-level factorial design, (ii) symmetrically arrayed star points located on the factor axes, and (iii) replicated center-points.

The CCC and CCF designs differ in how the star points, or axis points, are positioned. With the former design type, the axis points are located outside the factorial part of the composite design. This gives five levels for each investigated factor, which is advantageous in the estimation of a quadratic polynomial model. However, when it is desirable to keep the low and high factor levels unchanged, and still perform an RSM design, the CCF design is a viable alternative. In the CCF design, the axial points are centered on the faces of the cube or hypercube. This implies that all factors have three levels, rather than five, and that the experimental region is a cube or hypercube, and not a sphere. CCF is the recommended design choice for pilot plant and full scale investigations. Normally, RSM designs like CCC and CCF are used in conjunction with two to five factors. It is possible to explore as many as five factors in as few as 29 experiments. However, when moving up to six or more factors, there is a huge increase in the number of experiments, which means that RSM is less feasible in these cases.

Quiz II

Please answer the following questions:

Which tools are useful for evaluation of raw data?

Which tools are meaningful for diagnostic checking of a model?

When and why is it necessary to log-transform responses?

What measures had to be taken to improve the initial quadratic model of the truck engine data?

Which are the three actions that may occur after RSM?

What does interpolation mean? Extrapolation?

What does simplex optimization mean?

How does the optimizer make use of the existing regression model for starting simplex optimization?

What is of crucial importance regarding any final predicted best experiment?

Summary

In this chapter, we have discussed the experimental objective called optimization. For this purpose, we have reviewed the truck engine application, consisting of three factors, three responses, and a CCF design comprising 17 experiments. The problem formulation steps of this application were reviewed at the beginning of this chapter. The next part was devoted to the introduction of composite designs, and much emphasis was given to a geometrical interpretation of their properties. Following the introduction of the family of composite designs, the truck engine application was outlined in detail. The analysis of this data-set resulted in three excellent models, which were used for optimization. The optimization, in turn, leads on to the identification of an optimal factor combination at which the desired response profile was satisfied. The detected optimal point was later verified experimentally.

16 Experimental objective: Optimization (Level 3)

Objective

The objective of this chapter is to expand our discussion regarding optimization designs. Composite designs, which are commonly used in optimization, were outlined in Chapter 15. Now, we shall discuss alternative design families which also accomplish response surface modelling, and, in the end, optimization. We will discuss three additional design families, namely (i) three-level full factorial designs, (ii) Box-Behnken designs, and (iii) D-optimal designs. Our description of these families will focus on their geometrical properties and their applicability. A comparison with the composite family is also provided.

Three-level full factorial designs

Three-level full factorial designs are simple extensions of the two-level full factorials. Figure 16.1 displays the simplest design in this family, which is the 3^2 full factorial design. This notation is understood as 2 factors varied in three levels. We can see that the two factors are explored in a regular array consisting of nine experiments. The three-level full factorial design in three factors, denoted 3^3, is constructed in a similar manner. As seen in Figure 16.2, this design requires 27 experiments. The next design in this family is the 3^4 full factorial design, which requires as many as 81 experiments. Consequently, this design and other three-level factorials in more factors are rarely used, because the number of experiments becomes prohibitively large. In essence, this means that the 3^2 and 3^3 protocols are the ones to pay closest attention to. Of course, both these designs benefit from the addition of replicated center-points. Also observe that the 3^2 design is equivalent to the CCF design in two factors.
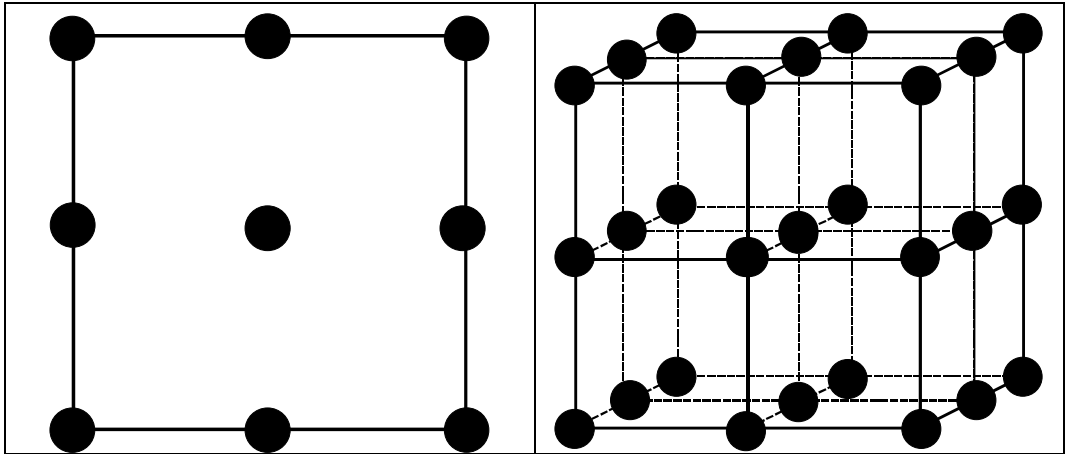
Figure 16.1: (left) Geometry of the 3^2 full factorial design.
Figure 16.2: (right) Geometry of the 3^3 full factorial design.

Box-Behnken designs

The Box-Behnken (BB) class of designs represents another family of designs employing only three levels per varied factor. These designs are economical in the number of runs required, as is exemplified by the three-factor version shown in Figure 16.3. An example worksheet of this design is shown in Figure 16.4. We can see that the experiments are located at the mid-point of each edge of the cube. An optional number of center-points are encoded by the solid dot in the interior part of the design. The BB-designs are useful if experimenting in the corners is undesirable or impossible. Mostly, BB-designs are used when investigating three or four factors. Observe that this design type is not defined for the two-factor situation.

	Worksheet						
	1	2	3	4	5	6	7
	Exp No	Exp Name	Run Order	Incl/Excl	x1	x2	x3
1	1	N1	15	Incl	−1	−1	0
2	2	N2	2	Incl	1	−1	0
3	3	N3	5	Incl	−1	1	0
4	4	N4	1	Incl	1	1	0
5	5	N5	10	Incl	−1	0	−1
6	6	N6	11	Incl	1	0	−1
7	7	N7	6	Incl	−1	0	1
8	8	N8	3	Incl	1	0	1
9	9	N9	14	Incl	0	−1	−1
10	10	N10	7	Incl	0	1	−1
11	11	N11	8	Incl	0	−1	1
12	12	N12	12	Incl	0	1	1
13	13	N13	13	Incl	0	0	0
14	14	N14	4	Incl	0	0	0
15	15	N15	9	Incl	0	0	0
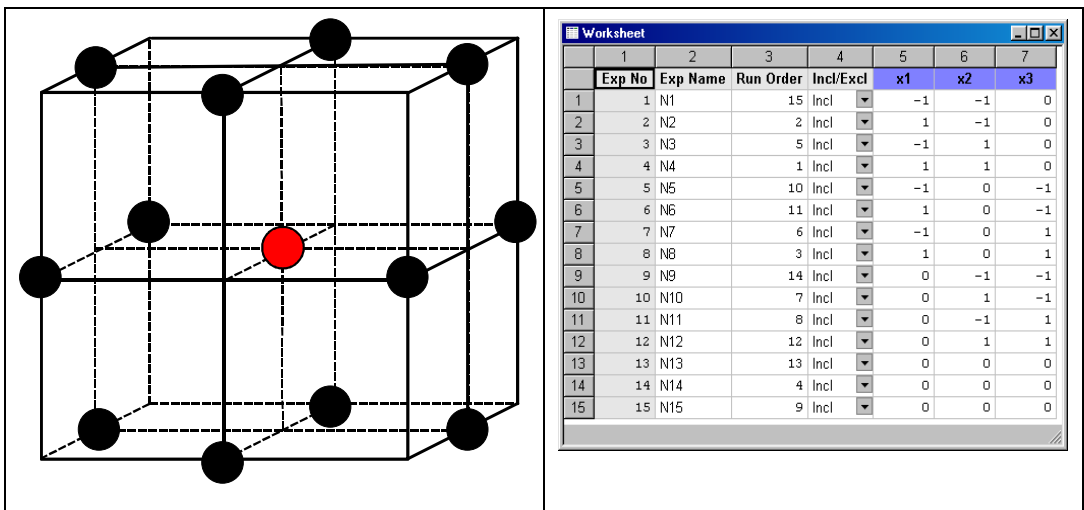
Figure 16.3: (left) The Box-Behnken design in three factors.
Figure 16.4: (right) Example worksheet of the design given in Figure 16.3.

A comparison of Composite, Three-level factorial, and Box-Behnken designs

We have now examined three families of classical optimization designs, which are useful when the experimental domain exhibits a regular geometry. Hence, it is of interest to compare them with regards to the number of runs required. The table displayed in Figure 16.5 provides an overview of the number of encoded experiments of composite, three-level full factorial, and Box-Behnken (BB) designs, for tasks involving two to five factors.

# Factors	CCC/CCF	Three-level	Box-Behnken
2	8 + 3	9 + 3	-----
3	14 + 3	27 + 3	12 + 3
4	24 + 3	81 + 3	24 + 3
5	26 + 3	243 + 3	40 + 3

Figure 16.5: *A comparison of composite, three-level factorial, and Box-Behnken designs.*

Overall, the CCC and CCF designs are most economical. Some parsimony is provided by the BB-designs in three and four factors as well, but with five factors the BB design is not an optimal choice. The main drawback of the three-level full factorial designs is the exponential increase in the number of experiments. Only for two and three factors are the number of experiments acceptable.

However, regardless of which design is favored, there is another factor lurking in the background, which may be the primary determinant for the choice of design. If, in screening, a two-level fractional factorial design had been used, and there is a desire to build on this design in a later stage, it is difficult to establish a three-level full factorial or Box-Behnken design. Rather, the designs from the composite family are more tractable in this respect. Hence, if screening is initiated with a two-level fractional factorial design, optimization is usually conducted with a composite design.

Properties of classical RSM designs

We shall now examine some requirements on, and properties of, these RSM designs. Good RSM designs have to (i) allow the estimation of the model parameters with low uncertainty, (ii) give small prediction errors, (iii) provide prediction errors independent of direction, (iv) allow a judgement of model adequacy, and (v) limit the number of runs.

The first consideration means that the confidence intervals of the model's regression coefficients must be as narrow as possible. The second and third criteria combined imply that a model should predict well, and that the error in the predictions should not be affected by the direction in factor space in which predictions are made. The prediction error should only be influenced by the distance to the design center. This property of equal prediction precision in all directions is normally referred to as *rotatability*. Some of the designs studied here are rotatable, and some are not. Rotatability may be controlled in composite designs by regulating the distance of the star points to the design center, but this is an advanced topic.

Further, the fourth consideration is that replicated experiments must be incorporated in the design, so that a lack of fit test can be carried out in the ANOVA procedure. This is ensured with a suitable number of replicated center-point experiments. The number of center-points

is also used for controlling the *orthogonality*, or *near-orthogonality*, of a design, which is another important property. Finally, we have the fifth consideration, which concerns the number of experiments. It is always desirable to reduce the number of experiments, but not by sacrificing the relevance or quality of the work. There is a trade-off between the sharpness of the information conveyed by a design and the number of runs it requires.

D-optimal designs

The last topic of this chapter concerns the family of D-optimal designs. D-optimal designs are computer generated designs that are valuable alternatives to classical RSM designs whenever the experimenter wants to create a non-standard design, or the experimental region is substantially distorted. The theory concerning the construction of D-optimal designs is an advanced topic, which is dealt with in Chapter 18. Here we focus on the practical aspects of applying D-optimal designs.

We envision six optimization situations, in which D-optimal designs are particularly useful. The first situation relates to an irregular experimental region. This occurs when there is a part of the experimental region in which we are unable or unwilling to do experiments. Here, classical designs of regular geometry are less applicable, and a D-optimal design of irregular geometry becomes more appropriate. An example is shown in Figure 16.6. The second situation deals with mixture design applications of irregular regions. Figure 16.7 provides an example of this. This situation may occur when there are lower and upper bounds, other than 0 and 1, of the mixture factors. Classical mixture designs of regular geometry are inapplicable in this case.
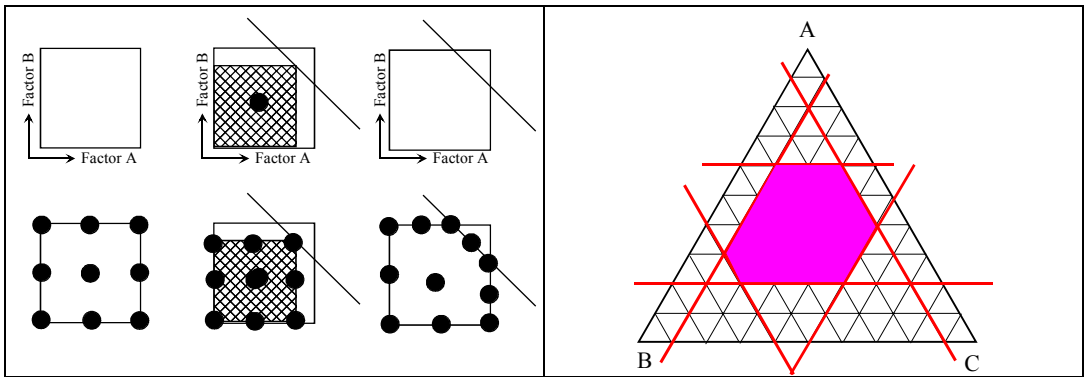
Figure 16.6: (left) When to use D-optimal design: Irregular experimental region.
Figure 16.7: (right) When to use D-optimal design: Mixture design of irregular region.

The third situation concerns the use of qualitative factors. Usually, the best level of a qualitative factor will have been ascertained already in the screening phase, and this level is then kept fixed throughout the optimization. Sometimes, however, one may need further study of a qualitative factor in optimization, to finally discover its best setting. In such circumstances, one may lay out a D-optimal design to reduce the number of runs required. As an example, consider the design shown in Figure 16.8, which involves a study of three quantitative factors and one qualitative factor at two levels. As seen, the D-optimal approach applies well in this situation, by encoding a 23 run protocol.
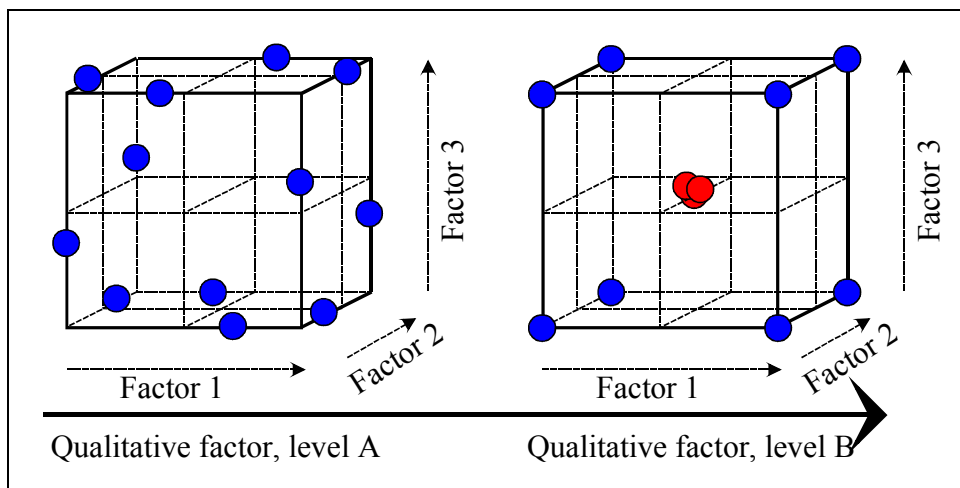
Figure 16.8: When to use D-optimal design: Qualitative factors at two levels.

The fourth optimization situation where a D-optimal design is valuable is in the treatment of more than six factors. A D-optimal design is capable of substantially reducing the number of runs required. Figure 16.9 summarizes the most plausible design options and their recommended number of experiments for optimization of 5, 6 and 7 factors. Note that all the listed designs advocate the use of three center-points. We can see that in all cases the D-optimal design is more parsimonious than any other alternative, except for the five factor situation where a CCC or CCF is just as economical.

#Factors	CCC/CCF	BB	D-opt
5	26 + 3	40 + 3	26 + 3
6	44 + 3	48 + 3	35 + 3
7	78 + 3	56 + 3	43 + 3

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{11}x_1^2 + b_{22}x_2^2 + b_{33}x_3^2 + b_{111}x_1^3 + e$$
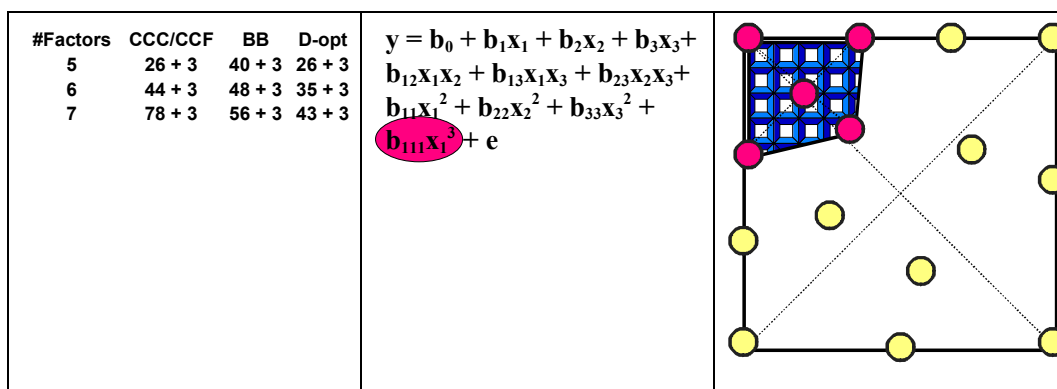
Figure 16.9: (left) When to use D-optimal design: Resources for fixed number of runs.
Figure 16.10: (middle) When to use D-optimal design: Fitting of special regression model.
Figure 16.11: (right) When to use D-optimal design: Inclusion of existing experiments.

We now continue with the fifth situation in which a D-optimal design is a good choice. This is when the experimenter wishes to fit a special regression model. Perhaps, for instance, a quadratic model in three factors must be updated with a single cubic model term, but an upgrading to the full cubic model is judged unwarranted. Figure 16.10 illustrates this kind of special model updating. Such a selective model updating has become increasingly common in DOE, because of the modest increase in the number of runs required. Finally, we have the sixth situation, which relates to the inclusion of existing experiments into a new design. This

might be relevant if we have a limited number of interesting experiments already done, which cover a small region, and we want to supplement these in the best possible manner. This is graphically illustrated in Figure 16.11.

In summary, a D-optimal design is useful in optimization when we want to create an "odd" design. Observe that such non-standard designs often exhibit irregular geometric properties and therefore might be a little more difficult to analyze than standard DOE-protocols.

Quiz

Please answer the following questions:

What is a three-level full factorial design?

Why is a three-level full factorial design an unrealistic choice for four and more factors?

What is a Box-Behnken design?

When is a BB-design attractive?

Which are the five requirements of good RSM designs?

When are D-optimal designs relevant in optimization?

How many factors can conveniently be optimized with a composite design?

How many factors can conveniently be optimized with a D-optimal design?

Summary

In this chapter, we have considered alternatives to composite designs for optimization. Thus, we have outlined in detail three-level full factorial designs and Box-Behnken designs, and compared these with the CCC and CCF designs. These design families are appropriate for exploring regular experimental domains. Should the experimental domain be irregular, designs drawn from the D-optimal family are more appropriate. We indicated six optimization situations in which the D-optimal methodology is of particular interest.

17 Experimental objective: Robustness testing (Level 2)

Objective

The objective of this chapter is to describe robustness testing, and this is done by adhering to the DOE framework outlined in Chapters 1-12. Our example regards the robustness assessment of a high-performance liquid chromatography (HPLC) analysis system, the kind of equipment routinely used in the analytical chemistry laboratory. This application involves the variation of five factors, four quantitative and one qualitative, and the measurement of three responses. Of these three responses, one is more important than the others and has to be robust towards changes in the five factors. The other two responses should also not change too much. Thus, the experimental objective is to verify the robustness of the important response. We will also highlight some statistical experimental designs which are well suited for the robustness testing experimental objective. In addition, the analysis of the HPLC data will show how a number of raw data evaluation tools, and regression modelling diagnostic tools, work on a real, complicated example. The chapter ends with a discussion of four limiting cases of robustness testing.

Introduction to robustness testing

The objective of robustness testing is to design a process, or a system, so that its performance remains satisfactory even when some influential factors are allowed to vary. In other words, what is wanted is to minimize our system's sensitivity to changes in certain critical factors. If this can be accomplished, it obviously provides many advantages, like easier process control, wider range of applicability of product, higher quality of product, and so on. A robustness test is usually carried out *before* the release of an almost finished product, or analytical system, as a last test to ensure quality. Such a design is usually centered around a factor combination which is currently used for running the analytical system, or the process. We call this point the *set point*. The set point might have been found through a screening design, an optimization design, or some other identification principle, such as, written quality documentation. The objective of robustness testing is, therefore, to explore robustness close to the set point.

Background to General Example 3

The example that we have chosen as an illustration originates from a pharmaceutical company. It represents a typical analytical chemistry problem within the pharmaceutical industry, and other industrial sectors. In analytical chemistry, the HPLC method is often used for routine analysis of complex mixtures. It is therefore important that such a system will work reliably for a long time, and be reasonably insensitive to varying chromatographic conditions. Some HPLC equipment is shown in Figure 17.1, and a typical chromatogram in Figure 17.2.
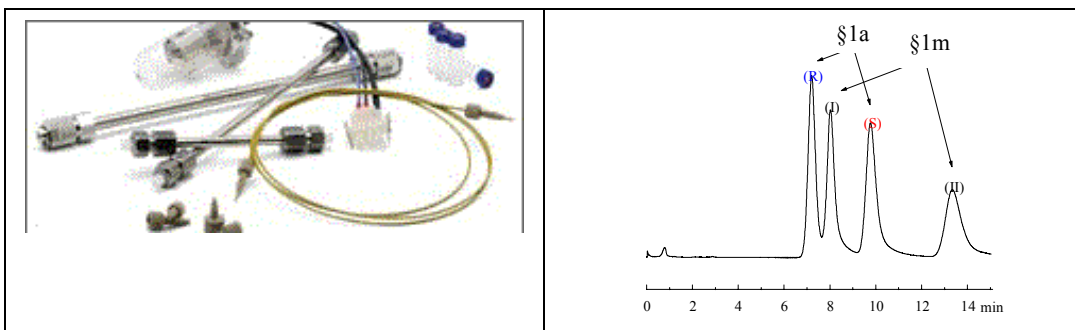
Figure 17.1: (left) Some HPLC equipment.
Figure 17.2: (right) Schematic of an HPLC chromatogram.

The investigators studied five factors, namely *amount of acetonitrile in the mobile phase*, *pH of mobile phase*, *temperature*, *amount of the OSA counter-ion in the mobile phase,* and *batch of stationary phase (column)*, and mapped their influence on the chromatographic behavior of two analyzed chemical substances. The low and high settings of these factors are found in Figure 17.3, together with the relevant measurement units. Observe that the last factor is of a qualitative nature. To study whether these factors had an influence on the chromatographic system, the researchers used a 12 run experimental design to encode 12 different chromatographic conditions. For each condition, three quantitative responses, reflecting the capacity factors of the two analytes (compounds) and the resolution between the analytes, were measured. These responses are summarized in Figure 17.4.

	Name	Abbr.	Units	Type	Use	Settings
1	**AcN**	Ac	%	Quantitative	Controlled	25 to 27
2	**pH**	pH		Quantitative	Controlled	3.8 to 4.2
3	**Temp**	Te	°C	Quantitative	Controlled	18 to 25
4	**OSA**	OS	mM	Quantitative	Controlled	0.09 to 0.11
5	**Column**	Co		Qualitative	Controlled	ColA,ColB

	Name	Abbr.	Units
1	**k1**	k1	
2	**k2**	k2	
3	**Res1**	Re1	

Figure 17.3: (left) Overview of factors studied in the HPLC application.
Figure 17:4: (right) Overview of measured responses in the HPLC application.

In chromatography, the objective is separation of the analytes, or, rather, separation in a reasonable time. To get separation, one must first have retention. Thus, the retention of each

analyte is important, and this response is given by the capacity factor, k. Furthermore, the degree of separation between two analytes is estimated as the resolution between two adjacent peaks in the chromatogram. A resolution of 1 is considered as the minimum value for separation between neighboring peaks, but for complete baseline separation a resolution of >1.5 is necessary. As the resolution value approaches zero, it becomes more difficult to discern separate peaks. The goal of this study was to constantly maintain a resolution of 1.5 or higher for all chromatographic conditions. No specific target values were given for the two capacity responses.

Problem Formulation (PF)

Selection of experimental objective

The problem formulation is important in DOE. The first step of the problem formulation is the selection of a relevant experimental objective. In the HPLC application, the chosen experimental objective was *robustness testing*. In robustness testing of an analytical equipment, the central idea is to lay out a condensed statistical design around the set point, and this design should have small changes in the factors. One then wants to find out how sensitive or insensitive the measured responses are to these small changes. Certain responses might be insensitive to such small changes in the factors and therefore claimed to be robust. Other responses might be sensitive to alterations in the levels of some factors, and these factors must then be better controlled. In the HPLC application, the set point was obtained from the written quality documentation of the analytical equipment. In the following, we will relate how the planning and execution of the robustness testing design was centered around this point.

Specification of factors

Robustness testing is useful for creating the final version of the quality documentation of an analytical system, in the sense that such a text should contain recommended factor settings between which robustness can be assured. Usually, the experimenter has a rough idea about such relevant factors settings for the system, and a robustness testing design is often used for critically corroborating these. In the HPLC case, five factors were selected for the robustness test. These factors are shown in Figures 17.5 – 17.9. The factor levels were chosen by considering which changes may occur during a normal day in the laboratory, that is, mild but controlled shifts. The first factor is amount of acetonitrile in the mobile phase, for which 25 and 27% were set as low and high levels. Acetonitrile is used as an organic eluent modifier. The second factor is the pH of the mobile phase, and this factor was varied between 3.8 and 4.2. The third factor is the laboratory temperature, which was investigated between 18 and 25°C. The fourth factor is the amount of counter-ion octanesulphonic acid, OSA, which is used for regulating a compound's retention. OSA was varied between 0.09 and 0.11 mM, and 0.01 is considered to be a large change. Finally, we have the fifth factor, the type of stationary phase (column), which display a qualitative, or discrete, nature. This factor represents the switch between two batches of stationary phase, for simplicity encoded as ColA and ColB.
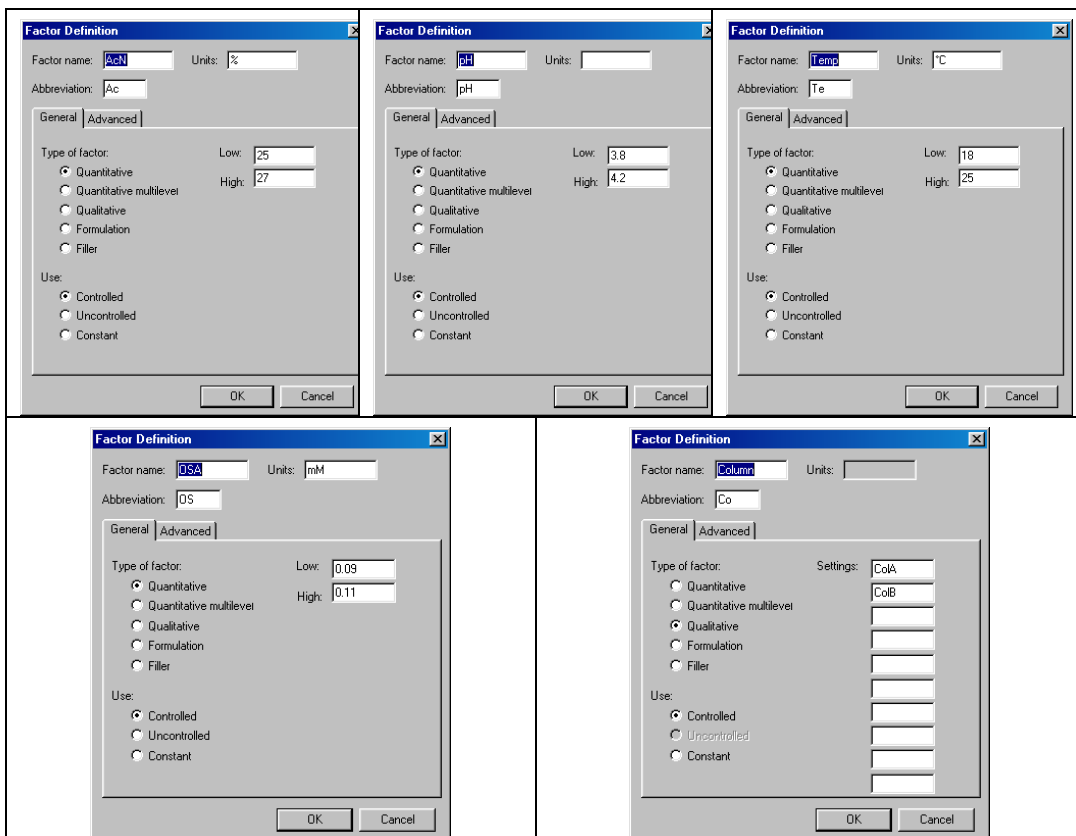
Figure 17.5: (upper left) Factor definition of amount of acetonitrile in mobile phase.
Figure 17.6: (upper middle) Factor definition of pH of mobile phase.
Figure 17.7: (upper right) Factor definition of temperature.
Figure 17.8: (lower left) Factor definition of OSA counter-ion in mobile phase.
Figure 17.9: (lower right) Factor definition of batch of stationary phase.

Specification of responses

The next step in the problem formulation concerns the specification of which responses to monitor. It is important that the selected responses reflect the important properties of the system studied. In HPLC, the resolution of the system, that is, the extent to which adjacent peaks in the resulting chromatogram can be separated from each other, is of paramount importance. A complete baseline separation corresponds to the ideal case, and this is inferred when the resolution exceeds 1.5. Therefore, the goal was to maintain a resolution of 1.5 or higher. The other type of response of interest is the retention, which corresponds to the time required for migration of a compound from the injector to the detector. A measure of the retention is given by the capacity factor. In this example, the goal was retention for a reasonable, short, time, but no specific targets were set. The three responses, denoted k_1 for capacity factor 1, k_2 for capacity factor 2, and Res1 for the resolution between compounds 1 and 2, are overviewed in Figures 17.10 – 17.12.
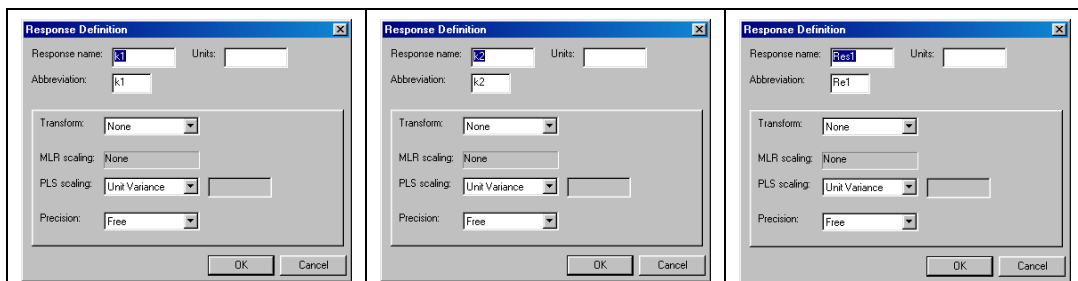
Figure 17.10: *(left) Response definition of capacity factor 1.*
Figure 17.11 *(middle) Response definition of capacity factor 2.*
Figure 17.12: *(right) Response definition of resolution between compounds 1 and 2.*

Selection of regression model

The fourth step of the problem formulation concerns the selection of an appropriate regression model. In principle, this selection can be made from either of three classes of models, namely linear, interaction or quadratic regression models. A linear model is an appropriate choice for robustness testing. One reason for this is that we are usually exploring each factor within a narrow range, and hence strong departures from linearity are unlikely. Also, we are only interested in identifying which factors should be better controlled, and an adequate answer to this will normally be provided by a linear model. The investigators choose a linear model in all factors for the HPLC-data.

Generation of design and creation of worksheet

The last two stages of the problem formulation deal with the generation of the statistical experimental design and the creation of the corresponding worksheet. We will treat these two steps at the same. Because the ideal result in a robustness testing study is identical response values for each trial, the selected experimental protocol may well be a low-resolution screening design. In particular, we think here of resolution III fractional factorial designs, and designs drawn from the Plackett-Burman family. Both these types of designs were introduced in Chapters 13 and 14.

The experimenters behind the HPLC data selected a 2^{5-2} fractional factorial design consisting of eight runs. This design supports a linear polynomial model, and the worksheet is provided in Figure 17.13. The five columns with numbers 5 – 9 denote the factors, and the three columns with numbers 10 – 12 the responses. The first eight rows correspond to the eight corner-experiments. It is seen that these eight runs are supplemented with four experiments of a "center-point-nature". They have numerical settings close to the center-point values in the four quantitative factors, but not in the qualitative factor. No factor level located half-way between ColA and ColB is definable. This means that the worksheet can be said to contain two pairs of replicates, one pair being allotted to each level of the qualitative factor.

Figure 17.13: Experimental data pertaining to the HPLC application.

Common designs in robustness testing

Resolution III fractional factorial designs and Plackett-Burman designs are two design types often used for robustness testing. We will now briefly review their properties, and a detailed explanation is retrievable from Chapters 13 and 14. A fractional factorial design is a design which consists of only a part, a fraction, of all the theoretically possible corner-experiments. As an example, consider Figure 17.14, which shows the 2^3 full factorial design of the CakeMix example. This 2^3 full factorial design may be split into two balanced half-fractions, one shown in Figure 17.15 and the other shown in Figure 17.16. These two half-fractions are encoded by two versions of the so called 2^{3-1} fractional factorial design, and imply that three factors can be investigated in four runs. From a design point of view, these two half-fractions are equivalent, but in reality one of them may be preferred to the other for practical reasons.
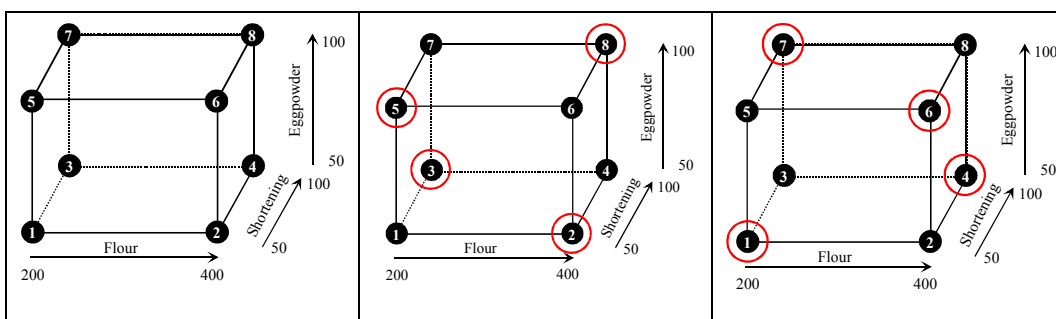
Figure 17.14. (left) Geometry of the 2^3 design.
Figure 17.15. (middle) Geometry of one version of the 2^{3-1} design.
Figure 17.16: (right) Geometry of the second version of the 2^{3-1} design.

In the HPLC application five factors are studied. With five factors in two levels, 32 corner-experiments are defined. These 32 corners may be divided into four quarter-fractions consisting of eight experiments each. The 2^{5-2} fractional factorial design used specifies the testing of one such quarter-fraction. This design has resolution III. The concept of design

resolution is closely linked to the kind of regression model supported by the design. Resolution III designs are the most reduced, that is, contain fewest experiments, and are therefore useful for estimating linear models. This makes them well suited to robustness testing. Resolution IV and V fractional factorial designs include more experiments, and may be used to calculate more elaborate regression models. This was explained in Chapter 13.

We now turn to the other common design family in robustness testing, the Plackett-Burman class of designs. Plackett-Burman designs are orthogonal two-level experimental designs, which can be used to fit linear models. We will now consider some of their features, by inspecting the 12 run PB-design in 11 factors given in Figure 17.17. In the construction of a PB-design one needs access to an appropriate starting row comprising minus one and plus one digits. Once the first row has been acquired the construction of a Plackett-Burman design proceeds as follows. The first row is cyclically permuted one-step to the right and shifted one-step down. This means that the entries of the first row will appear in the same order in the second row, but be shifted to the right. Next, the third row is generated by shifting the second row one-step to the right, and so on. This continues until the next shifting would have resulted in the initial row, whereupon the permutation is terminated. In the current case, this means that row 11 is the last one to be created through the permutation procedure. Finally, a row of only minus one elements is added to the bottom of the design, followed by the insertion of the replicated center-points.

	1	5	6	7	8	9	10	11	12	13	14	15
	Exp No	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11
1	1	1	−1	1	−1	−1	−1	1	1	1	−1	1
2	2	1	1	−1	1	−1	−1	−1	1	1	1	−1
3	3	−1	1	1	−1	1	−1	−1	−1	1	1	1
4	4	1	−1	1	1	−1	1	−1	−1	−1	1	1
5	5	1	1	−1	1	1	−1	1	−1	−1	−1	1
6	6	1	1	1	−1	1	1	−1	1	−1	−1	−1
7	7	−1	1	1	1	−1	1	1	−1	1	−1	−1
8	8	−1	−1	1	1	1	−1	1	1	−1	1	−1
9	9	−1	−1	−1	1	1	1	−1	1	1	−1	1
10	10	1	−1	−1	−1	1	1	1	−1	1	1	−1
11	11	−1	1	−1	−1	−1	1	1	1	−1	1	1
12	12	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1	−1
13	13	0	0	0	0	0	0	0	0	0	0	0
14	14	0	0	0	0	0	0	0	0	0	0	0
15	15	0	0	0	0	0	0	0	0	0	0	0

Figure 17.17: Plackett-Burman design in 12 + 3 runs.

One should observe that the number of runs in a PB-design is a multiple of 4, which implies that the PB-designs nicely complement the fractional factorials family as their number of runs is a power of 2. In principle, this means that a fractional factorial design is selected whenever a robustness test is carried out in 8 or 16 runs, and a PB-design is selected when there is a need to use either 12 or 20 experimental trials. This concludes the introductory treatment of the robustness testing objective. We will now take a short quiz break, and then continue with the analysis of the HPLC-data.

Quiz

Please answer the following questions:

What is the typical objective in a robustness test?

Which type of regression model is appropriate for robustness testing?

Why is a low-resolution fractional factorial design appropriate?

Which numbers of runs are most suitable for a Plackett-Burman design? A fractional factorial design?

HPLC application

Evaluation of raw data

We start the evaluation of the raw data by inspecting the replicate error of each response. Three relevant plots of replications are displayed in Figures 17.18 – 17.20.
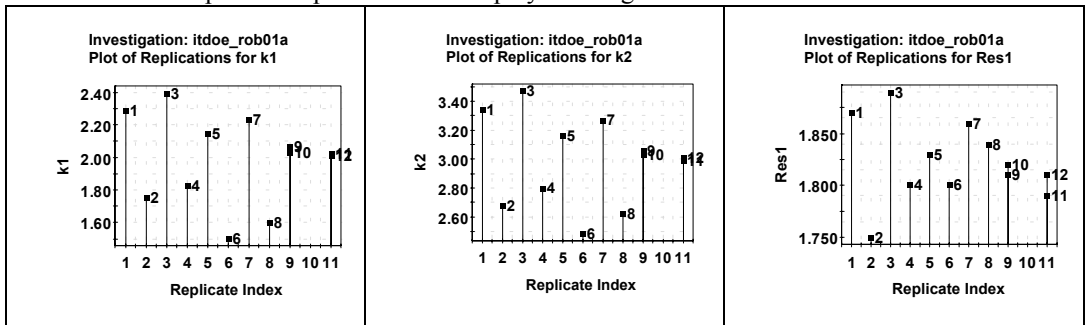
Figure 17.18: *(left) Replication plot of capacity factor 1.*
Figure 17.19: *(middle) Replication plot of capacity factor 2.*
Figure 17.20: *(right) Replication plot of resolution.*

As seen, the replicate error is small for each response, which is what was expected. We would not anticipate large drifts among the replicates, as we have deliberately set up a design where each run should ideally produce equivalent results. The numerical variation in the resolution response is small (Figure 17.20). The lowest measured resolution is 1.75 and the highest 1.89. Since the operative goal was to maintain the resolution above 1.5, we see even in the raw data that this goal is fulfilled, and this means that Res1 is robust.

Further, in the evaluation of the raw data, it is compulsory to check the data distribution of the responses, to reveal any need for response transformation. We may check this by plotting a histogram of each response. Such histograms are plotted in Figures 17.21 – 17.23, and they inform us that it is pertinent to work in the untransformed metric of each response.
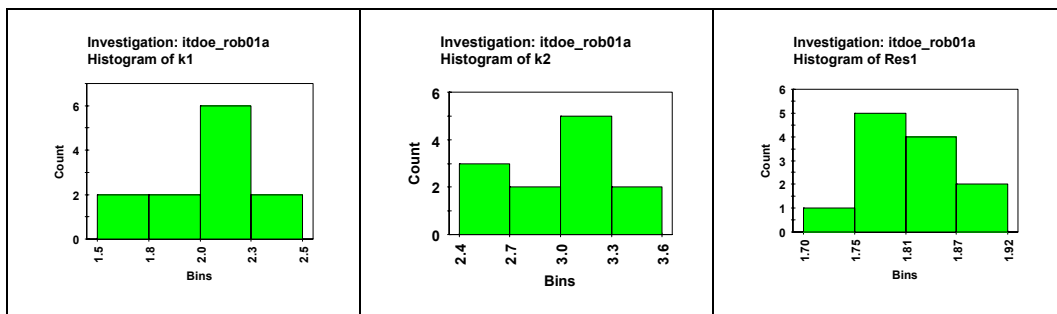
Figure 17.21: *(left) Histogram of capacity factor 1.*
Figure 17.22: *(middle) Histogram of capacity factor 2.*
Figure 17.23: *(right) Histogram of resolution.*

Another aspect of the raw data properties that might be worth considering is the condition number of the fractional factorial design. We observe in Figure 17.24 that the condition number is approximately 1.2, indicating a pertinent design. Furthermore, the correlation matrix, listed in Figure 17.25, indicates that the three responses are strongly correlated. Hence, the fitting of three regression (MLR) models with identical composition of model terms is reasonable.
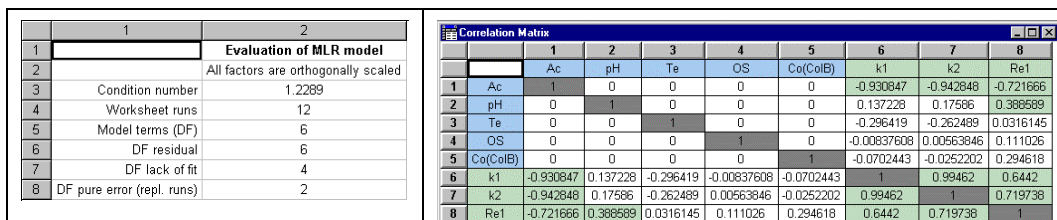
Figure 17.24: *(left) Evaluation of design condition number.*
Figure 17.25: *(right) Plot of correlation matrix revealing the strong inter-relatedness among the three responses.*

Regression analysis

The regression analysis phase in robustness testing is carried out in a manner similar to that of screening and optimization. However, here the focus is primarily on the R^2 and Q^2 parameters, and on the analysis of variance results, but not so much on residual plots and other graphical tools. The reason for this is that the interest in robustness testing lies in classifying the regression model as significant or not significant. With such information it is then possible to understand the robustness. Another modelling difference with respect to screening and optimization is that model refinement is not carried out.

We fitted a linear model with 6 terms to each response. The overall results of the model fitting is displayed in Figure 17.26. It is seen that the predictive power ranges from poor to excellent. The Q^2 values are 0.92, 0.96, and 0.12, for k_1, k_2, and Res1, respectively. In robustness testing the ideal result is a Q^2 near zero. Hence, the Q^2 of 0.12 for Res1 is an indication of an extremely weak relationship between the factors and the response, that is, it seems that the response is robust. The low Q^2 for Res1 might be explained by the fact that this response is almost constant across the entire design, and hence there is little response variation to account for. The high Q^2's for k_1 and k_2, on the other hand, indicate that these

responses are sensitive to the small factor changes. However, for these latter responses, there is no point in making any robustness statement, as there are no specifications to be met.

k1	DF	SS	MS (variance)	F	p	SD
Total	12	48.287	4.024			
Constant	1	47.471	47.471			
Total Corrected	11	0.816	0.074			0.272
Regression	5	0.798	0.16	53.785	6.66E-05	0.4
Residual	6	0.018	2.97E-03			0.054
Lack of Fit (Model Error)	4	0.017	4.23E-03	9.447	0.098	0.065
Pure Error (Replicate Error)	2	8.95E-04	4.48E-04			0.021

N = 12 Q2 = 0.915 CondNo = 1.2289
DF = 6 R2 = 0.9782 Y-miss = 0
R2Adj = 0.96 RSD = 0.0545
ConfLev = 0.95

k2	DF	SS	MS (variance)	F	p	SD
Total	12	108.479	9.04			
Constant	1	107.485	107.485			
Total Corrected	11	0.994	0.09			0.301
Regression	5	0.983	0.197	112.622	7.60E-06	0.444
Residual	6	0.01	1.75E-03			0.042
Lack of Fit (Model Error)	4	0.01	2.47E-03	8.126	0.113	0.05
Pure Error (Replicate Error)	2	6.07E-04	3.04E-04			0.017

N = 12 Q2 = 0.9594 CondNo = 1.2289
DF = 6 R2 = 0.9895 Y-miss = 0
R2Adj = 0.9807 RSD = 0.0418

Res1	DF	SS	MS (variance)	F	p	SD
Total	12	39.874	3.323			
Constant	1	39.858	39.858			
Total Corrected	11	0.016	1.48E-03			0.038
Regression	5	0.013	2.51E-03	4.066	0.059	0.05
Residual	6	3.70E-03	6.17E-04			0.025
Lack of Fit (Model Error)	4	3.45E-03	8.63E-04	6.901	0.131	0.029
Pure Error (Replicate Error)	2	2.50E-04	1.25E-04			0.011

N = 12 Q2 = 0.1215 CondNo = 1.2289
DF = 6 R2 = 0.7719 Y-miss = 0
R2Adj = 0.5819 RSD = 0.0248

Figure 17.26: *(upper left) Summary of fit plot for the HPLC data.*
Figure 17.27: *(upper right) ANOVA of capacity factor 1.*
Figure 17.28: *(lower left) ANOVA of capacity factor 2.*
Figure 17.29: *(lower right) ANOVA of resolution.*

The results of the second diagnostic tool, the analysis of variance, are summarized in Figures 17.27 – 17.29. Remembering that the upper p-value should be smaller than 0.05 and the lower p-value larger than 0.05, we realize that the former test is a borderline case with respect to Res1, because the upper listed p-value is 0.059. This suggests that the model for Res1 is not significant, and therefore that Res1 is robust.

This concludes the regression analysis phase. The derived models will now be used in a general discussion concerning various outcomes of robustness testing.

First limiting case – Inside specification/Significant model

The first limiting case is *inside specification and significant model*. The HPLC application contains one example of this limiting case, the Res1 response. We know from the initial raw data assessment that this response is robust, because all the measured values are inside the specification, that is, above 1.5. Actually, as seen in Figure 17.30, the measured values are all above 1.75. The question of a significant model, however, is more debatable. It is possible to interpret the obtained regression model as a weakly significant regression equation. We will do so in this section for the sake of illustration. The classification of the model as significant is based on a joint assessment of the low, but positive, Q^2, seen in

Figure 17.31, and the significant linear term of acetonitrile, seen in Figure 17.32. Hence, Res1 may be regarded as an illustration of the first limiting case.
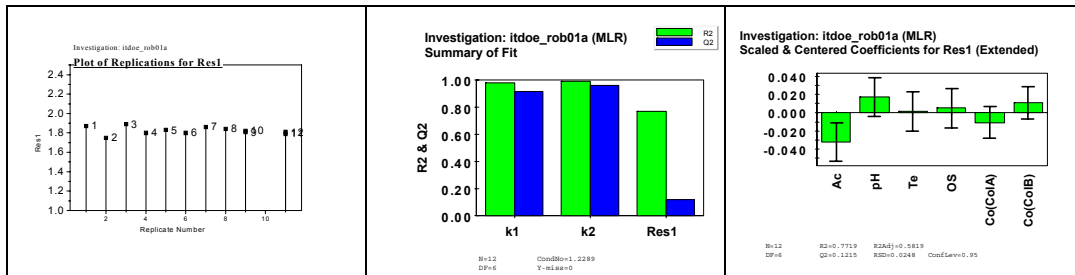
Figure 17:30: *(left) Plot of replications for Res1.*
Figure 17.31: *(middle) Plot of R^2 and Q^2.*
Figure 17.32: *(right) Regression coefficients of model for Res1.*

An interesting consequence of these modelling results is that it appears possible to relax the factor tolerances and still maintain a robust system. For instance, the model interpretation reveals that the amount of acetonitrile might be allowed to be as high as 28%, without compromising the goal of maintaining a resolution above 1.5.

Furthermore, in robustness testing it may also be of interest to estimate the response values of the most extreme experiments. Figure 17.32 gives guidance of how to obtain such estimates. We can see that one extreme experimental condition is given by the factor combination low AcN, high pH, high Temp, high OSA, and ColB, and the other extreme experiment by the reversed factor pattern. Figure 17.33 gives these Res1 predictions, and as seen they are both valid with regard to the given specification.

Figure 17.33: *Extreme cases predictions for the Res1 response.*

Second limiting case – Inside specification/Non-significant model

The second limiting case is *inside specification and non-significant model*. This is the ideal outcome of a robustness test. Again, we may use the Res1 response as an illustration. We know that the measured values of this response are all inside specification, and it is also possible to interpret the regression model obtained as non-significant. This classification of the model as non-significant is contrary to the classification made in the previous section, but is still reasonable and is made for the purpose of illustrating the second limiting case.

In general, to assess model significance two diagnostic tools emerge as more appropriate than others. The first tool consists of the R^2 and Q^2 parameters. When these are simultaneously near zero, as is the situation in Figure 17.34, we have the ideal case. This

means that we are trying to model a system in which there is *no* relationship among the factors and the response in question. In reality, however, one has to expect slight deviations from this outcome. A typical result is the case when R^2 is rather large, in the range of 0.5-0.8, and Q^2 low or close to zero. As seen in Figure 17.35 this is the case for Res1, which points to an insignificant model.
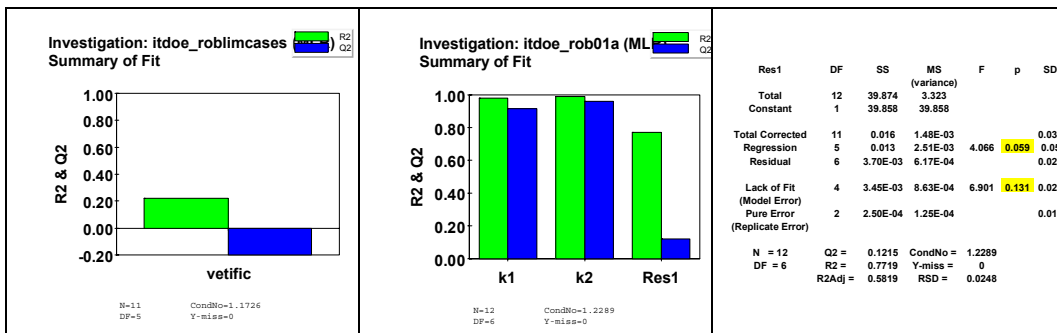
Figure 17.34: (left) Schematic of a situation of simultaneously low R^2 and Q^2.
Figure 17.35: (middle) Plot of R^2 and Q^2.
Figure 17.36: (right) ANOVA of Res1.

The second important modelling tool relates to the analysis of variance, and particularly the upper F-test, which is a significance test of the regression model. We can see in Figure 17.36 that the Res1 model is weakly insignificant, because the p-value of 0.059 exceeds 0.05. Hence, we conclude that no useful model is obtainable. When no model is obtainable it is reasonable to anticipate that all the variation in the experiments can be interpreted as a variation around the mean. This variation can then be seen as the mean value +/- t-value * standard deviation.

Third limiting case – Outside specification/Significant model

The third limiting case is *outside specification and significant model*. This limiting case occurs whenever a significant regression model is calculated, and the raw response data themselves do not meet the goals of the problem formulation. We will use the second response, k_2, of the HPLC data to illustrate this limiting case. In order to accomplish a meaningful illustration, we will have to define a specification for k_2, for example that k_2 should be between 2.7 and 3.3. This kind of specification for a capacity factor is uncommon in the pharmaceutical industry, but one is set here for the sake of illustration.

We start by assessing the statistical behavior of the k_2 regression model. This behavior is evident from Figure 17.37, which indicates the sensitivity to small factor changes of k_2 (as well as k_1). In order to understand what is causing this susceptibility to changes in the factors, it is necessary to consult the regression coefficients displayed in Figure 17.38.
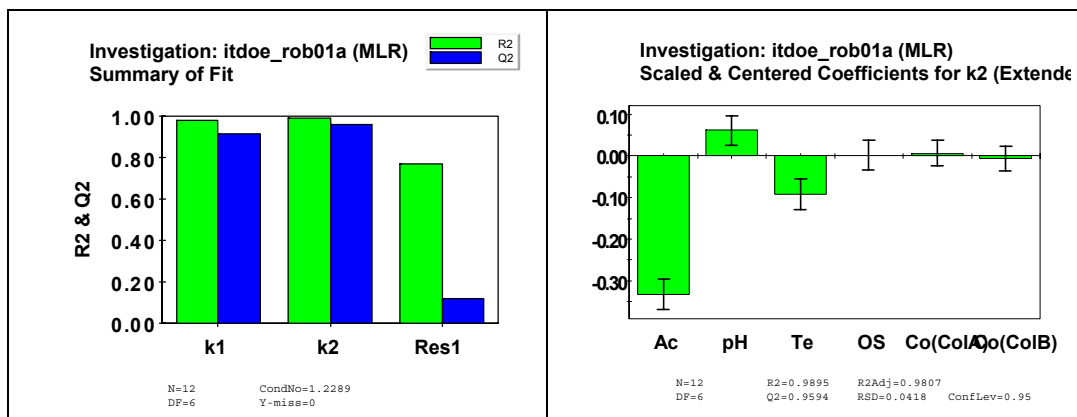
Figure 17.37: *(left) Plot of R^2 and Q^2.*
Figure 17.38: *(right) Regression coefficients of model for k_2.*

We can see that it is mainly acetonitrile, pH and temperature, which affect k_2. By using the procedure which was outlined in connection with the first limiting case, we can change the factor intervals to accomplish two things, namely (i) get k_2 inside specification and (ii) produce a non-significant model, that is, approach the second limiting case. First of all, it is possible to predict the most extreme experimental values (in the investigated area) of k_2. These are the predictions listed on the first two rows in Figure 17.39, and they amount to 2.50 and 3.49. Clearly, we are outside the 2.7 – 3.3 specification.

Figure 17.39: *Predictions of k_2 corresponding to interesting factor combinations.*

In order to get within the specification, we must adjust the factor ranges of the three influential factors, and this is illustrated in rows three and four (Figure 17.39). To get a non-significant regression model even narrower factor intervals are needed. This is done as follows: The regression coefficient of acetonitrile is –0.33 and its 95% confidence interval ±0.036. These numbers mean that this coefficient must be decreased by a factor of 10, that is, be smaller than around –0.03, in order to make this factor non-influential for k_2. Since this coefficient corresponds to the response change when the amount of acetonitrile is increased by 1%, from 26 to 27%, we realize that the new high level must be lowered from 27 to 26.1%. A similar reasoning applies to the new lower factor level. Hence, the narrower, more robust, factor tolerances of acetonitrile ought to be between 25.9 and 26.1%. A similar

reasoning for the temperature factor indicates that the factor interval should be decreased to one-third of the original size. Appropriate low and high levels thus appear to be 20 and 23 degrees centigrade. The obtained predictions are listed in rows five and six of Figure 17.39.

This concludes our treatment of the third limiting case, and our message has been that it is possible to use the modelling results for understanding how to reformulate the factor settings so that robustness can be obtained.

Fourth limiting case – Outside specification/Non-significant model

The fourth limiting case is *outside specification and non-significant model*. This limiting case may be the result when the derived regression model is poor, and there are anomalies in the data. It is important to uncover such anomalies, because their presence will influence the modelling. An informative graphical tool for understanding whether this is happening is the replicate plot. Figure 17.40 shows an example in which one strong outlier is present, which will invalidate the robustness. Figure 17.41 depicts a case where all the replicated center-points have much higher response values than the other runs. This pattern hints at curvature and implies non-robustness. A third common situation, which partly resembles the first case, is when one experiment deviates from the rest and also falls outside some predefined robustness limits. This is shown in Figure 17.42.
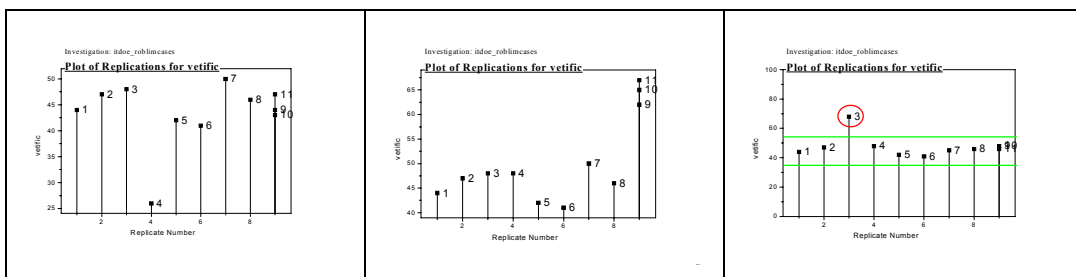
Figure 17.40: (left) Replicate plot – One strong outlier.
Figure 17.41: (middle) Replicate plot – Center-points with higher values than other points.
Figure 17.42: (right) Replicate plot – One point outside specifications (horizontal lines).

Evidently, there can be several underlying explanations for this limiting case, and we have just mentioned a few. Therefore, we consider this limiting case as the most complex one. In summary, we have now described four limiting cases of robustness testing, but it is important to realize that robustness testing results are not limited to these four extremes. In principle, there is a gradual transition from one limiting case to another, and hence an infinite number of outcomes are conceivable.

Summary of the HPLC application

The application of DOE in the robustness testing of the HPLC system was very successful. With this approach it was possible to infer the robustness of the Res1 response. From a tutorial point of view, the HPLC application is good for several reasons. It represents a realistic case in which all the necessary steps for verifying the robustness of an analytical system are illustrated. Furthermore, this data set also allowed us to study how tools for evaluation of raw data, tools for regression analysis, and tools for model use, were put into

practice in a real situation. It should be clear from this application, that the modelling step is of crucial importance in robustness testing, as it is linked to an understanding of the nature of the robustness or non-robustness. We also used the HPLC study for discussing four limiting cases of robustness testing.

Quiz II

Please answer the following questions:

Which tools are useful for evaluation of raw data?

Which tools are useful for diagnostic checking of a regression model?

Which two evaluation criteria may be used for understanding the results of robustness testing?

Which are the four limiting cases of robustness testing?

Summary

In this chapter, we have discussed the experimental objective called robustness testing. For this purpose, we have reviewed the HPLC application, consisting of five factors, three responses, and a 2^{5-2} fractional factorial design in 12 experiments. The problem formulation steps of this application were reviewed at the beginning of the chapter. The next part of the chapter was devoted to the discussion of common designs in robustness testing. Following this introduction, the HPLC application was outlined in detail. The analysis of this data-set resulted in excellent models for two responses and a weak model for one response. These models and the raw experimental data were used to establish the robustness of the Res1 response.

Part III

Chapters 18-24

Additional Topics

18 D-optimal design (Level 3)

Objective

This chapter is devoted to an introduction of D-optimal design. Our aim is to describe the D-optimality approach and to point out when to use it. In the generation of a D-optimal design, the selection of experimental runs can be driven towards fulfilling different criteria. For this purpose, we will explain two common evaluation criteria, one called the G-efficiency and the other the condition number. We will describe a number of applications where D-optimal design is useful.

D-optimal design

When to use D-optimal design

D-optimal designs are used in screening and optimization, as soon as the researcher needs to create a non-standard design. The D-optimal approach can be used for the following types of problem:

- irregular experimental regions (Figures 18.1 – 18.3),
- multi-level qualitative factors in screening (Figure 18.4)
- optimization designs with qualitative factors (Figure 18.5)
- when the desired number of runs is smaller than required by a classical design (Figures 18.6 – 18.7)
- model updating (Figure 18.8 – 18.9)
- inclusions of already performed experiments (Figures 18.10 – 18.11)
- combined design with process and mixture factors in the same experimental plan (Figure 18.12)
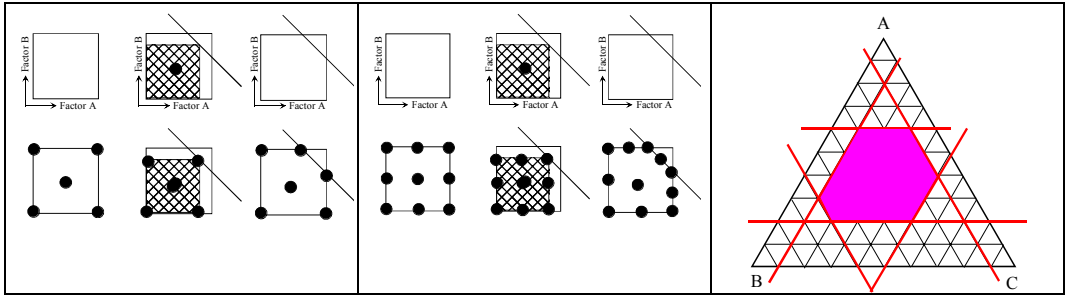
Figure 18.1: *(left) Irregular experimental region – screening.*
Figure 18.2: *(middle) Irregular experimental region – optimization.*
Figure 18.3: *(right) Irregular experimental region – mixture design.*
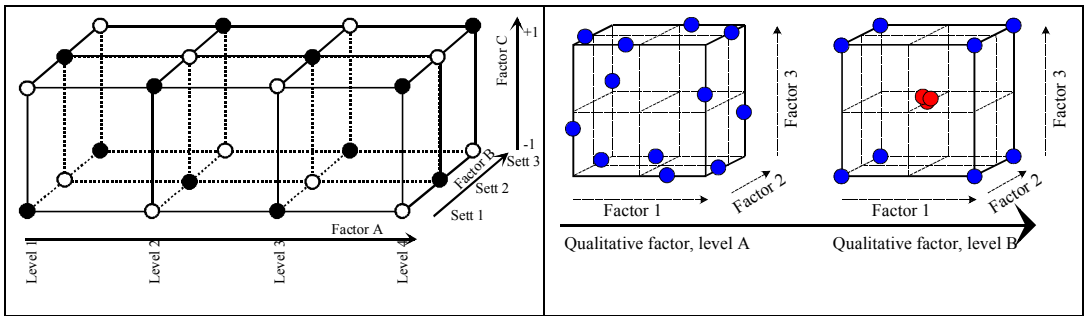
Figure 18.4: *(left) Multi-level qualitative factors in screening.*
Figure 18.5: *(right) Qualitative factors in optimization.*

# Runs	# Center-points	# Total runs	Design type
8	3	11	Frac Fac
11	3	14	D-optimal
12	3	15	PB
16	3	19	Frac Fac

#Factors	CCC/CCF	BB	D-opt
5	26 + 3	40 + 3	26 + 3
6	44 + 3	48 + 3	35 + 3
7	78 + 3	56 + 3	43 + 3

Figure 18.6: *(left) D-optimal design when number of runs is smaller than that for a classical design – screening.*
Figure 18.7: *(right) D-optimal design when number of runs is smaller than that for a classical design – optimization.*

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{22}x_2^2 + e$$

$$b_{11}x_1^2 \quad b_{33}x_3^2$$

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{11}x_1^2 + b_{22}x_2^2 + b_{33}x_3^2 + b_{111}x_1^3 + e$$

Figure 18.8: *(left) Model updating – screening.*
Figure 18.9: *(right) Model updating – optimization.*
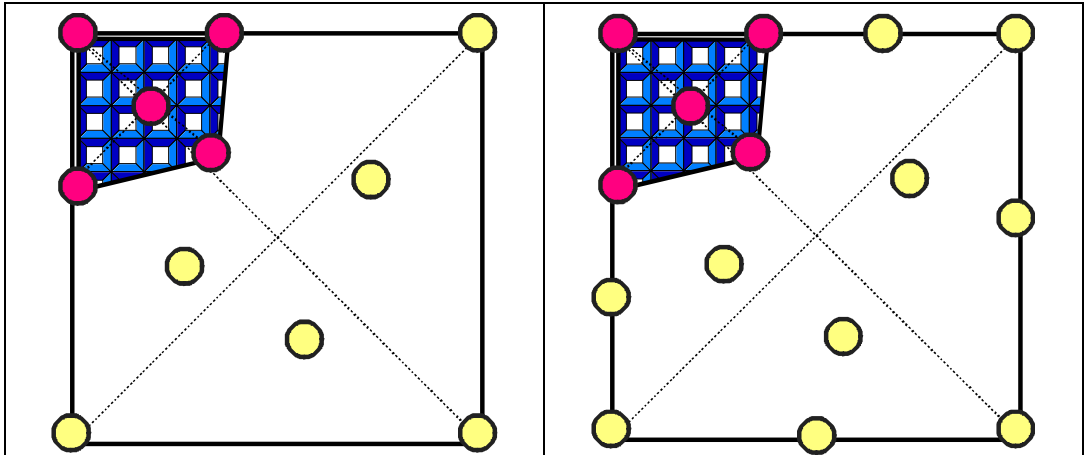
Figure 18.10: (left) Inclusion of already performed experiments – screening.
Figure 18.11: (right) Inclusion of already performed experiments – optimization.
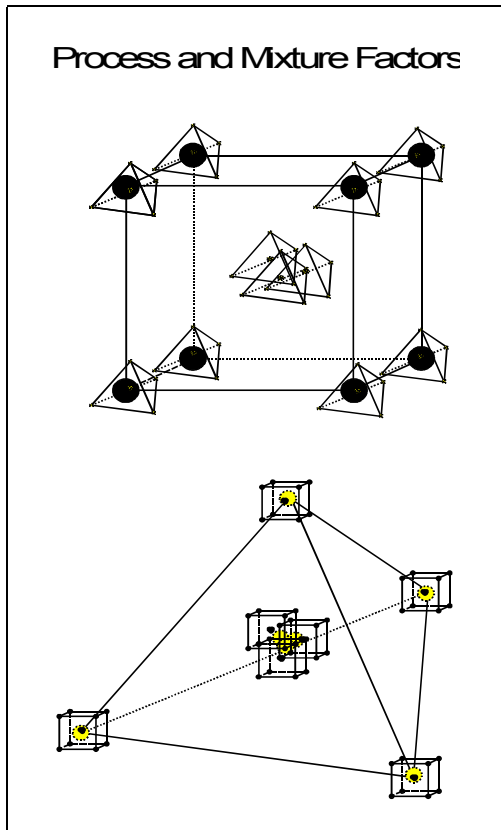
Figure 18.12: Designs involving process and mixture factors at the same time.

We will now give an explanation of D-optimal design, and thereafter use D-optimal design to illustrate three of these problem types, namely model updating, multi-level qualitative factors, and combined designs of process and mixture factors.

An introduction

A D-optimal design is a computer generated design, which consists of the *best* subset of experiments selected from a candidate set. The candidate set is the pool of theoretically possible and practically conceivable experiments. To understand what is the *best* design, one must evaluate the selection of experimental runs according to a given criterion. The criterion used here is that the selected design should maximize the determinant of the matrix X'X for a given regression model. This maximization criterion also explains how the letter "D" in D-optimal was derived, from D in determinant. Thus, the D-optimal approach means that N experimental runs are chosen from the candidate set, such that the N trials maximize the determinant of X'X. Equivalently, we can say that the N runs span the largest volume possible in the experimental region. This search for the best subset of experiments is carried out using an automatic search algorithm in MODDE.

In the following, we will present two ways of illustrating the principles underlying D-optimal design. The first example addresses D-optimality from a least squares analysis perspective. The second example is a geometric representation of what we hope to accomplish with a D-optimal selection of experiments.

An algebraic approach to D-optimality

We will now address D-optimality from an algebraic perspective, which is useful for the least squares analysis. Consider Figure 18.13, which shows the design matrix of the 2^2 full factorial design. We call the two factors x_1 and x_2. As seen in Figure 18.14, this design supports an interaction model in the two factors. This model, which is listed in equation form in Figure 18.14, may be rewritten in matrix form, as shown in Figure 18.15. This matrix expression hints at how to solve the problem algebraically, and estimate the model parameters, contained in the b-vector of regression coefficients. It can be shown (but not here) that the best set of coefficients according to least squares is given by the equation $b = (X'X)^{-1}X'y$.

run	x_1	x_2	Model	Model in matrix form
1	-1	-1	$y = b_0 + b_1x_1 + b_2x_2 +$	$y = Xb + e$
2	1	-1	$+ b_{12}x_1x_2 + e$	➔ $b = (X'X)^{-1}X'y$
3	-1	1		
4	1	1		

Figure 18.13: (left) Design matrix of the 2^2 design.
Figure 18.14: (middle) Model supported by design shown in Figure 18.13.
Figure 18.15: (right) Supported model in matrix notation.

One crucial step in the estimation of the regression coefficients is the computation of the inverse of the X'X matrix, $(X'X)^{-1}$, and the properties of this matrix are intimately linked to the D-optimality approach. In order to compute the regression coefficients, we must extend the design matrix of Figure 18.13 to become the computational matrix, which is depicted in

two versions in Figures 18.16 and 18.17. Figure 18.16 displays which column of the computational matrix is used for the estimation of which regression coefficient, and Figure 18.17 shows a condensed computational matrix devoid of the surrounding information. The matrix of Figure 18.17 is equivalent to the matrix X in Figure 18.15.

run	const.	x1	x2	x1x2
1	1	-1	-1	1
2	1	1	-1	-1
3	1	-1	1	-1
4	1	1	1	1
	b0?	b1?	b2?	b12?

X

$$\begin{pmatrix} 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

X'

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 \end{pmatrix}$$

Figure 18.16: *(left) Alignment of columns and estimated coefficients.*
Figure 18.17: *(middle) The X-matrix.*
Figure 18.18: *(right) The X'-matrix.*

Figure 18.18 illustrates the transpose of X, that is X', which is needed to compute the X'X matrix displayed in Figure 18.19. Evidently, X'X is a symmetrical matrix with the element four in all entries of the diagonal. The value of 4 arises because the design has four corner-point experiments. When inverting this matrix, the matrix of Figure 18.20 is the result. The properties of this inverse matrix strongly influence the precision of the estimated regression coefficients. Figure 18.21 reveals that three terms are used in the calculation of confidence intervals for coefficients, namely, (i) the inverse of X'X, (ii) the residual standard deviation of the model, and (iii) the Student's t parameter. From this it follows that the smallest $(X'X)^{-1}$, or, alternatively, the largest X'X is beneficial for the precision of the regression coefficients.

(X'X)

$$\begin{pmatrix} 4 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix}$$

$(X'X)^{-1}$

$$\begin{pmatrix} 0.25 & 0 & 0 & 0 \\ 0 & 0.25 & 0 & 0 \\ 0 & 0 & 0.25 & 0 \\ 0 & 0 & 0 & 0.25 \end{pmatrix}$$

Precision in b from:

$(X'X)^{-1}$ * RSD * t

➔ smallest $(X'X)^{-1}$

➔ largest X'X

Figure 18.19: *(left) The X'X-matrix.*
Figure 18.20: *(middle) The $(X'X)^{-1}$-matrix.*
Figure 18.21: *(right) How the properties of $(X'X)^{-1}$ influence regression coefficient estimation.*

A geometric approach to D-optimality

We will now use another simple example to illustrate the D-optimal approach from a geometrical perspective. Imagine that our experimental objective is to uncover the effects of two factors, x_1 and x_2, on a response y. Furthermore, both x_1 and x_2 are investigated at three levels, denoted −1, 0 and +1. Then, as seen in Figure 18.22, we have a candidate set comprising nine experiments. Our proposed regression model is that $y = b_0 + b_1x_1 + b_2x_2 +$ e. This model contains only three parameters and hence only three experiments are needed. For the sake of simplicity, let us also assume that we wish to do only three experiments. In such a case the question that we have to ask ourselves is *how are we going to select these N = 3 runs from the nine-run candidate set*? Observe that there are (9! / (3!*6!)) = 84 ways of selecting 3 trials out of 9.
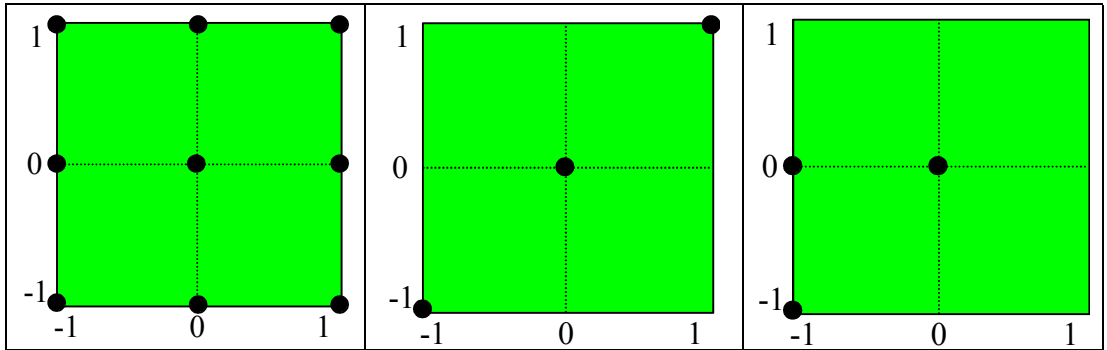
Figure 18.22: (left) Nine-run candidate set.
Figure 18.23: (middle) Three-run selection with determinant 0
Figure 18.24: (right) Three-run selection with determinant 1.

We will employ the D-optimal approach and choose the three runs that maximize the determinant of the relevant X'X matrix. And, for comparative purposes, we provide four supplementary choices of experiments, which are *not* D-optimal. Figures 18.23 – 18.26 show some alternative choices of three experiments. The three runs selected in Figure 18.23 encode an X'X matrix of determinant 0, that is, the experiments span no volume, or, as here, area, of the experimental region. Next, we have Figure 18.24, which illustrates a positioning of experiments corresponding to a determinant of 1. Similarly, Figure 18.25 corresponds to a determinant of 4, Figure 18.26 a determinant of 9, and Figure 18.27 a determinant of 16. We can see that the selection in Figure 18.27 has the best coverage of the experimental region. It also exhibits the largest determinant, and would therefore provide the most precise estimates of the coefficients of the proposed regression model $y = b_0 + b_1 x_1 + b_2 x_2 + e$. Observe that any selection of three corners out of four, would result in a determinant of 16.
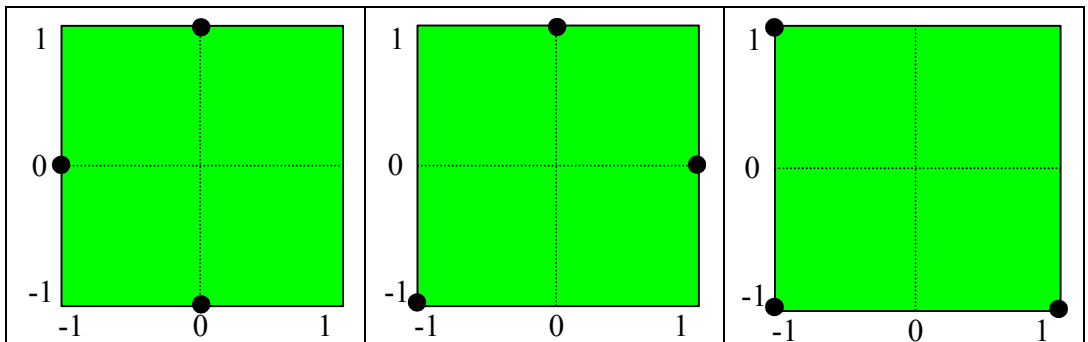
Figure 18.25: (left) Three-run selection with determinant 4.
Figure 18.26: (middle) Three-run selection with determinant 9.
Figure 18.27: (right) Three-run selection with determinant 16.

How to compute a determinant

We will now review how to compute a determinant. As an illustration, we shall consider the selection of three experiments corresponding to a determinant of 4. The distribution of these three experiments are portrayed in Figure 18.28.
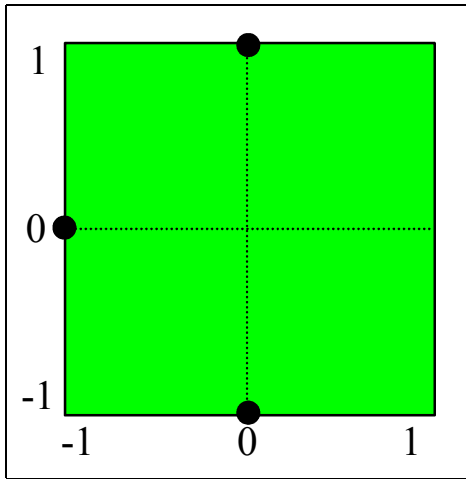
Figure 18.28: *Three-run selection with determinant 4.*

The three selected experiments represent the matrix X shown in Figure 18.29, which, in turn, may be transposed into the matrix given in Figure 18.30. The resulting matrix, X'X, is listed in Figure 18.31, and it is for this matrix that we wish to calculate the determinant.

1	-1	0	1	1	1	3	-1	0
1	0	-1	-1	0	0	-1	1	0
1	0	1	0	-1	1	0	0	2

Figure 18.29: *(left) Matrix X corresponding to example in Figure 18.28.*
Figure 18.30: *(middle) Matrix X' corresponding to example in Figure 18.28.*
Figure 18.31: *(right) Matrix X'X corresponding to example in Figure 18.28.*

In order to compute the determinant of X'X one proceeds as is schematically illustrated in Figures 18.32 and 18.33, and arithmetically outlined in Figure 18.34. Figures 18.32 and 18.33 display a slightly elongated version of X'X. Firstly, the elements of three successive diagonals going from top left to bottom right are multiplied diagonal-wise, and then summed. Secondly, the elements of three successive diagonals running from top right to bottom left are multiplied diagonal-wise, and then summed. Finally, as shown in Figure 18.34, the second sum is then subtracted from the first. The result is the determinant of the X'X matrix.

3	-1	0	3	-1		3	-1	0	3	-1
-1	1	0	-1	1		-1	1	0	-1	1
0	0	2	0	0		0	0	2	0	0

(3*1*2) + (-1*0*0) + (0*-1*0) – (0*1*0) – (0*0*3) – (2*-1*-1) = 4

Figure 18.32: *(upper left) Computation of determinant partial sum.*
Figure 18.33: *(upper right) Computation of determinant partial sum.*
Figure 18.34: *(lower part) Computation of determinant.*

Features of the D-optimal approach

The D-optimal approach assumes that the selected regression model is "correct" and "true". This means that experiments will be selected that are maximally suited for the identified model. Usually, such experiments are positioned on the boundary of the experimental region. As a consequence, the D-optimal approach is sensitive to the choice of model. The D-optimal algorithm requires the number of design points to equal the number of parameters of the model, but often the selected number of design runs is greater than the number of parameters. Further, from a user point of view, it is important to know that some of the weaknesses of the D-optimal approach are counterbalanced by the software. For instance, the concept of lack of fit is not recognized by the D-optimal criterion, and to compensate for this MODDE automatically appends three replicates at the overall center-point. Also, the software mitigates against the sensitivity to the selected model by considering *potential terms*. A potential term is a higher order term not included in the original model, but which may be used if necessary. The selected design contains experiments which will support the estimation of potential terms.

Evaluation criteria

The D-optimal selection of experiments may be evaluated by means of several criteria. Here, we will describe two criteria, termed the *condition number* and the *G-efficiency*. The condition number is a measure of the sphericity and symmetry of a D-optimal design. Recall that this diagnostic tool was introduced in Chapter 10. Formally, the condition number is the ratio of the largest and smallest singular values of the X-matrix. Informally, it is roughly equivalent to the ratio of the largest and smallest design diagonals. For an orthogonal design the condition number is 1, and the higher the number the less orthogonality. Thus, if the D-optimal approach is directed towards this evaluation criterion, designs are proposed in which the experiments are positioned as orthogonally as possible. The second evaluation criterion, the G-efficiency, compares the efficiency, or performance, of a D-optimal design to that of a fractional factorial design. G-efficiency is computed as $Geff = 100*p/n*d$, where p is the number of model terms, n is the number of runs in the design, and d the maximum relative prediction variance across the candidate set. The upper limit of Geff is 100%, which implies that the fractional factorial design was returned by the D-optimal search. We recommend a G-efficiency above 60-70%.

Model updating

We will now describe three applications of D-optimal designs and start with model updating. Model updating is common after screening, when it is necessary to unconfound two-factor interactions. Consider the laser welding application described in Chapter 13. The four factors varied in this application are listed in Figure 18.35. In the first instance, the researcher conducted a 2^{4-1} fractional factorial design with three center-points, that is, eleven experiments. In the analysis it was found that one two-factor interaction, the one between Po and Sp, was influential. However, because of the low resolution of this design, this two-factor interaction was confounded with another two-factor interaction, namely No*Ro. An escape-route out of this problem was to carry out the fold-over design, enabling resolution of Po*Sp from No*Ro, as well as resolution of the remaining four two-factor interactions.

Figure 18.35: *The four factors of the laser welding application.*

The disadvantage of making the fold-over was a lot of extra experiments. Eleven additional runs were necessary. An alternative approach in this case, less costly in terms of experiments, would be to make a D-optimal design updating, adding only a limited number of extra runs. In theory, only two extra experiments are needed to resolve the Po*Sp and No*Ro two-factor interactions. In practice, however, four additional runs plus one or two center-points to test that the system is stable over time, is a better supplementation principle. Such an updated design is listed in Figure 18.36. This design has condition number 1.4 and G-efficiency 71.9%. Many other D-optimal proposals with similar performance measures exist, and it is up to the user to select a preferred version for the given application.

Figure 18.36: *Worksheet from model updating by D-optimal principle.*

Multi-level qualitative factors

Degrees of freedom

In the second example of this chapter, our aim is to demonstrate that multi-level qualitative factors are well handled by D-optimal design. We shall here consider the Cotton study (more details are given in Chapter 20), where two multi-level qualitative factors, the cotton Variety and the cultivation Center, are explored at four and seven levels respectively. In screening, one way to address these factors is with the full 4*7 factorial design in 28 runs, but this is overly many experiments. Since in screening a linear model is usually adequate, a

design with around 15 runs would suffice. An estimate of how many runs might be appropriate can be deduced from the number of degrees of freedom (DF). In this case, a linear model would at least require 1 DF for the constant, 6 DF for the linear term of Center, and 3 DF for the linear term of Variety. This adds up to 10 DF. How to calculate these degrees of freedom is explained in Chapter 20. In addition, as a precautionary measure, we recommend that 5 extra DF be added to facilitate the regression modelling. This is how we ended up with the estimate of 15 DF, that is, that the design should contain around 15 runs. As part of the next step, the D-optimal algorithm is then instructed to search for the best possible design centered around the use of 15 runs.

Evaluation of alternative designs

The D-optimal algorithm was instructed to search for designs with 14, 15 and 16 runs, and for each number of runs 5 alternative proposals were formulated. Thus, in total, MODDE generated 15 D-optimal designs. It turned out that all five 14 run designs had the same G-efficiency, 57.1%. Similarly, among the 15 run designs all five had a G-efficiency of 54.5%, and among the 16 run designs the result was always 60.2%. Consequently, the ranking according to G-efficiency is that a 16 run design is the best choice, followed by the 14 and 15 run counterparts. The identical trend holds true when the condition number is used as the evaluation criterion. However, in addition to this technical judgement, it is often favorable to obtain a geometric appraisal of the situation. In particular, we think here of the concept of *balancing*. Balancing implies that each level of a qualitative factor is explored with the same number of runs. This is a feature which is not absolutely necessary, but is desirable since it makes the design easier to understand. Balancing may be checked manually.

In the case of varying number of levels among qualitative factors, balancing is typically sought for the qualitative factor with most levels. We will now explore the balancing in the designs generated with 14 and 16 runs. Figures 18.37 and 18.38 display the arrangement of experiments of two alternative 14 run proposals. With 14 runs in the design, each level of Center is tested twice, and each level of Variety either three of four times.

Figure 18.37: *(left) Arrangement of experiments of 14 run D-optimal design.*
Figure 18.38: *(right) Alternative appearance of 14 run design.*

In an analogous fashion, Figures 18.39 and 18.40 display two alternative 16 run designs. Here, the balancing takes place with the four-level qualitative factor in focus. Apparently, each level of Variety is investigated using four experiments and each level of Center in either two or three experiments.

Figure 18.39: *(left) Arrangement of experiments of 16 run D-optimal design.*
Figure 18.40: *(right) Alternative appearance of 16 run design.*

In conclusion, in the Cotton application we would recommend the use of one of the 14 run designs, as such a design is balanced with respect to the seven-level factor Center. This kind of design is comparable to any one of the D-optimal designs comprising 16 experiments. It is also recommended that two replicated experiments be included, and these should be positioned in different rows and columns. According to Figure 18.37, a possible combination of replicates might be V1/C1 and V2/C2, but not the combination V1/C1 and V1/C3.

Combining process and mixture factors in the same design

Now, as a third example, our intention is to outline how process and mixture factors can be combined in one design. For this purpose, we will study an optimization of bubble formation. This application has a little commercial value, but is instructive as it reflects one of the most complex types of problem to which D-optimal design is applied.

Children like to blow bubbles, but dislike bubbles which burst too rapidly. We decided to use the mixture design approach for investigating which factors may affect bubble formation and lifetime. We browsed the Internet looking for a suitable bubble mixture recipe, which we could use as a starting reference mixture. Then this recipe was modified using mixture design, and bubbles were blown for each new mixture. The four ingredients, mixture factors, are overviewed in Figure 18.41. These are dish-washing liquid 1 (Skona, ICA, Sweden), dish-washing liquid 2 (Neutral, ADACO, Sweden), tap water (Umeå, Sweden) and glycerol (Apotekets, Sweden).

	Name	Abbr.	Units	Type	Use	Settings
1	**Temp**	Te		Quantitative	Controlled	7 to 21
2	**Time**	Ti		Quantitative	Controlled	1 to 25
3	**DWL1**	DW1	Fraction	Formulation	Controlled	0 to 0.4
4	**DWL2**	DW2	Fraction	Formulation	Controlled	0 to 0.4
5	**Water**	Wa	Fraction	Formulation	Controlled	0.4 to 0.9
6	**Glycerol**	Gly	Fraction	Formulation	Controlled	0 to 0.2

Figure 18.41: *The six factors of the bubble formation example.*

We can see that both dish-washing liquids are allowed to vary between 0 and 0.4. In addition to these restrictions, a relational constraint was specified stating that the sum of the dish-washing liquids ought to be between 0.2 and 0.5. This extra constraint is given in Figure 18.42.

	1	2	3	4	5	6	7	8	9
		Temp	**Time**	**DWL1**	**DWL2**	**Water**	**Glycerol**	**Operator**	**Limit**
2	Exclude :			1	1			>	0.5
3	Exclude :			1	1			<	0.2

Figure 18.42: *The relational constraint imposed on the two dish-washing liquid factors.*

Besides the four mixture factors, the influences of two "process" factors, storage temperature and mixture settling time, were investigated. As seen in Figure 18.43, the measured response was the lifetime of bubbles produced with a child's bubble wand. The time until bursting was measured for bubbles of 4 – 5 cm diameter.

	Name	Abbr.	Units	Transform
1	**Lifetime**	Li		Log

Figure 18.43: *The measured response in the bubble formation example.*

Bubble application

How many runs are necessary?

The experimental objective selected was screening and the model was one with linear and interaction terms. Interactions were allowed among the process factors, between the process and mixture factors, but not among the mixture factors themselves. In order to understand the complexity of the problem, we may look at the schematic in Figure 18.44. We may view this problem as one based on a two-level factorial design, on which is superimposed an array of mixture tetrahedrons. Observe that in the drawing the tetrahedral structures are depicted as symmetrical, whereas in reality they are mutilated.
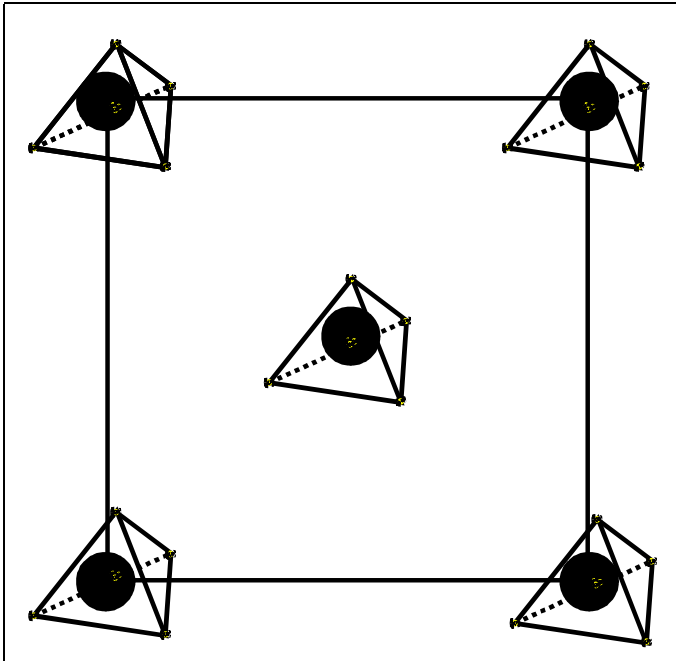
Figure 18.44: *Geometric representation of the bubble formation problem.*

In this case, the candidate set contains 169 experimental points, allocated as 48 extreme vertices, 48 edge points, 72 centroids of high-dimensional surfaces, and the overall centroid. This is a small candidate set and it was left intact by us. With six factors and a screening objective, the lead number of experiments N = 20 was suggested by MODDE. In the computation of this lead number, the number of degrees of freedom (DF) of the proposed model was taken into account. The necessary DF are calculated as follows: (a) 1 DF for the constant term, (b) 2 DF for the linear process terms, (c) 3 DF for the linear mixture terms, (d) 1 DF for the process*process interaction, and (e) 6 (2*3) DF for the process*mixture interactions. This makes 13 DF, and by adding 5 extra experiments to help the regression analysis we end up with N = 18. In addition, two supplementary runs were added to handle the additional complexity introduced by the linear constraint, thus giving N = 20 as the final lead number of experiments. Note that no replicates are included in this N = 20 estimate.

Generation of alternative D-optimal designs

In the generation of a D-optimal design an element of randomness is normally incorporated. In principle, this means that the D-optimal search is initiated from different starting points within the experimental region. Each selection may reach a local optimum, and the element of randomness means that one may get slightly different design proposals with the same number of runs.

Furthermore, the success of the D-optimality search also depends on the size of N. The recommended procedure would be to identify a lead number of design runs, N (here 20), selected by considering the number of model parameters plus some extra degrees of freedom, and then generate designs with N ± 4 runs and, say, five alternative versions for each level of N ± 4 runs. Adhering to this procedure, we generated 45 alternative D-optimal designs ranging from N = 16 to N = 24 and with five versions for each N. The G-

efficiencies of these designs are given in Figure 18.45. The best resulting design was one with N = 16 showing a G-efficiency of 76.1% and a condition number of 2.7. Its summary statistics are given in Figure 18.46, and they indicate a very good design.
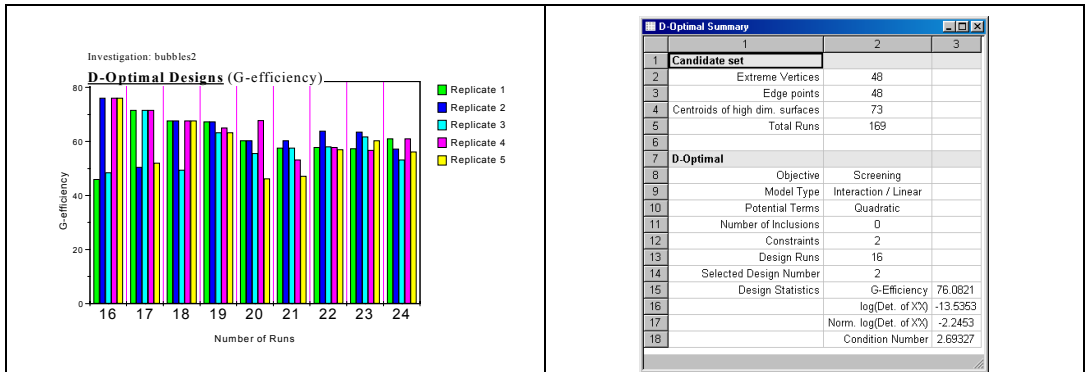
Figure 18.45: *(left) Evaluation of proposed D-optimal designs (G-efficiency).*
Figure 18.46: *(right) Summary statistics of selected D-optimal design.*

The experimental results

To the 16 experiments, we added two series of four replicates, making the entire screening design comprise 16 + 8 = 24 experiments. The resulting worksheet is listed in Figure 18.47. In retrospect, it may seem that overly many replicates were carried out, but at the time of experimenting we were unsure of the experimental reproducibility obtainable in a non-hi-tech kitchen. Since we wanted to obtain good insight into this source of uncertainty, we conducted 2*4 replicates. In other situations resembling ours, we would recommend 2*2 replicates.

Figure 18.47: *Experimental worksheet of bubble formation example.*

When we consider the resulting worksheet, we should remember its very special feature of simultaneously encoding changes in both process and mixture factors. It can be seen that the measured lifetime values vary between 11 seconds and 362 seconds (6.02 min), that is, a span of 1.5 orders of magnitude. Undoubtedly, this large difference must be attributed to the use of a sound D-optimal design. When the bubble-data obtained were further analyzed with regression analysis, a direction in which the bubble lifetime would be improved was discovered. Subsequently, a new D-optimal *optimization* design was laid out, and with this design a bubble lifetime as high as 22.28 min was scored. The key to prolonging the lifetime was to substantially increase the amount of glycerol.

Based on these bubble experiments, it was possible to conclude that the sequential strategy of first carrying out a screening study, followed by an optimization study, was very fruitful. Also, the D-optimal approach worked excellently for combining mixture and process factors.

Quiz

Please answer the following questions:

When is a D-optimal design applicable?

What is the meaning of the "D" in the name D-optimal?

Which property of the X'X matrix does the determinant reflect?

What is the basic assumption of the D-optimal approach concerning the selected model?

What is a potential term?

What is the condition number?

What is the G-efficiency?

What is the recommended procedure for establishing a D-optimal design?

Summary

In this chapter, we have provided an introduction to D-optimal design. Initially, the algebraic and geometrical aspects of the D-optimality criterion were introduced. We described how to use two criteria, the condition number and the G-efficiency, for evaluating a D-optimal search. Our treatment of D-optimal designs provided an overview of when D-optimal design is applicable, and three of these situations were discussed more thoroughly. These cases were (i) model updating, (ii) multi-level qualitative factors, and (iii) combined design of process and mixture factors.

19 Mixture design (Level 3)

Objective

Many commercially available chemical products are manufactured and sold as mixtures. The word mixture in this context designates a blend in which all ingredients sum to 100%. Examples of such mixtures are pharmaceuticals, gasolines, plastics, paints, and many types of food and dairy products. In most mixtures, the number of major components, sometimes also called constituents or ingredients, can vary between 3 and 8. Because mixtures may contain large numbers of components, it is necessary to use mixture design for finding mixture compositions with improved properties. Mixture design is a special topic within DOE, and many of the basic DOE concepts and modelling principles discussed so far, are applicable for design of mixtures. However, there are also some fundamental differences. It is the objective of this chapter to provide a short introduction to mixture design.

A working strategy for mixture design

The characteristic feature of a mixture is that the sum of all its ingredients is 100%. Firstly, this means that these components, mixture factors, cannot be manipulated completely independently of one another. Secondly, it means that their proportions must all lie somewhere between 0 and 1. As an example, we shall consider a tablet manufacturing application. Three constituents, microcrystalline cellulose, lactose and dicalciumphosphate-dihydrate, were mixed according to a ten-run mixture design. These factors will be referred to as *cellulose*, *lactose*, and *phosphate*. For each tablet produced, the release rate of the active substance was monitored. In the sections to come, we shall use this tablet example as an illustration of a proposed working strategy for mixture design. This strategy is influenced by the general DOE framework put forward in Part I of this course book.

Step 1: Definition of factors and responses

The first step of the mixture design strategy corresponds to the *definition of factors and responses*. For each factor low and high levels must be defined, as with regular process designs. In the context of mixture design, the term *bound* is used more frequently than *level*, and the lower and upper bounds of each mixture factor are given as proportions. As seen in Figure 19.1, all three mixture factors have the lower bound equal to 0 and the upper bound equal to 1. It is important to check the consistency of these bounds, because some combinations of bounds might be incompatible. For instance, proposing the alternative bounds cellulose (0/1), lactose (0/1) and phosphate (0.1/1) would not work in reality. The adjustments cellulose (0/0.9) and lactose (0/0.9) are necessary to make this mixture system realistic. The detection and fixing of inconsistent bounds is automatically taken care of by

MODDE. Further, as part of the first step, the responses of interest have to be specified. We can see from Figure 19.2 that the response of interest is the release rate of the active substance, which is to be maximized.

Figure 19.1: *(left) Factors of tablet manufacturing example.*
Figure 19.2: *(right) Response of tablet manufacturing example.*

Step 2: Selection of experimental objective and mixture model

The second step is the selection of experimental objective and mixture model. We have previously learnt that the choice of experimental objective, that is, screening, optimization, or robustness testing, has a decisive impact on the number of experimental trials. The experimental objective selected here was optimization. This objective needs the quadratic regression model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{33} x_3^2 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \varepsilon$$

with the overall external constraint $x_1 + x_2 + x_3 = 1$, and where x_k are the fractions of the three mixture ingredients, expressed in the 0 - 1 range.

The overall mixture constraint introduces a closure which produces some problems for the regression analysis, and some mathematical reparametrization is needed to alleviate the impact of this constraint. One approach is to calculate the regression coefficients with constraints imposed, linking certain groups of model terms to each other and hence making the individual terms inseparable. This is very similar to the situation occurring with expanded terms of multi-level qualitative factors (see Chapter 20). For instance, with the above model, it would be impossible to delete only the x_3^2 term. A removal of x_3^2 also requires that the two cross-terms $x_1 x_3$ and $x_2 x_3$ be removed from the model. Thus, these three terms should be treated as a unit, and be deleted from, or incorporated in, the model together.

Step 3: Formation of the candidate set

In the third step of the strategy, the candidate set is compiled. This is the pool of theoretically possible experiments from which is drawn a subset of experiments comprising the actual mixture design. Unwanted experiments may be deleted from the candidate set prior to the generation of the design. In the current example, the candidate set is small, because the mixture region is *regular* in geometry. This makes it easy to compute the extreme vertices, the centers of edges, the overall centroid, and so on, which build up the candidate set. Here, the extreme vertices are the extreme points of the experimental domain, and the overall centroid has properties reminiscent of a center-point experiment.

In the tablet formulation example, the candidate set comprises 3 extreme vertices, 3 centers of edges, 3 interior points and 1 overall centroid. However, in more complicated mixture applications, with several factors and elaborate constraints, the experimental region might be highly irregular. For such an application, the candidate set is usually larger in size and

also difficult to compile. Typically, the candidate set for an irregular mixture region consists of (i) the extreme vertices, (ii) the centers of edges, (iii) the centers of high-dimensional surfaces, and (iv) the overall centroid. From the resulting candidate set several alternative designs may be proposed D-optimally, each one corresponding to a unique selection of experiments.

Step 4: Generation of mixture design

In the generation of the experimental design, which is the fourth step of the working strategy, attention has to be paid to the specifications set in the three preceding steps. By considering these specifications, one finds that the classical *simplex centroid* design is applicable in the tablet example. This experimental protocol is shown in Figure 19.3.

	Exp No	Exp Name	Run Order	Incl/Excl	cellulose	lactose	phosphate	release
1	1	N1	10	Incl	1	0	0	197
2	2	N2	7	Incl	0	1	0	110
3	3	N3	4	Incl	0	0	1	324
4	4	N4	9	Incl	0.5	0.5	0	67
5	5	N5	2	Incl	0.5	0	0.5	362
6	6	N6	6	Incl	0	0.5	0.5	312
7	7	N7	1	Incl	0.666667	0.166667	0.166667	206
8	8	N8	3	Incl	0.166667	0.666667	0.166667	171
9	9	N9	8	Incl	0.166667	0.166667	0.666667	344
10	10	N10	5	Incl	0.333333	0.333333	0.333333	214

Figure 19.3: Experimental protocol of tablet manufacturing example.

We will now examine the geometrical properties of this design. With three mixture factors ranging between 0 and 1, the experimental region is a regular simplex. Such a simplex is illustrated in Figure 19.4. In this simplex, the orientation is as follows: Each vertex corresponds to a pure component, that is, only cellulose, only lactose, or only phosphate. To each vertex a component axis is associated, and at the common intersection of these axes the overall centroid of composition 1/3,1/3,1/3 is located. In order to well map this experimental region it is essential that the experiments are spread as evenly as possible. Figure 19.5 indicates the distribution of experiments of the selected design.
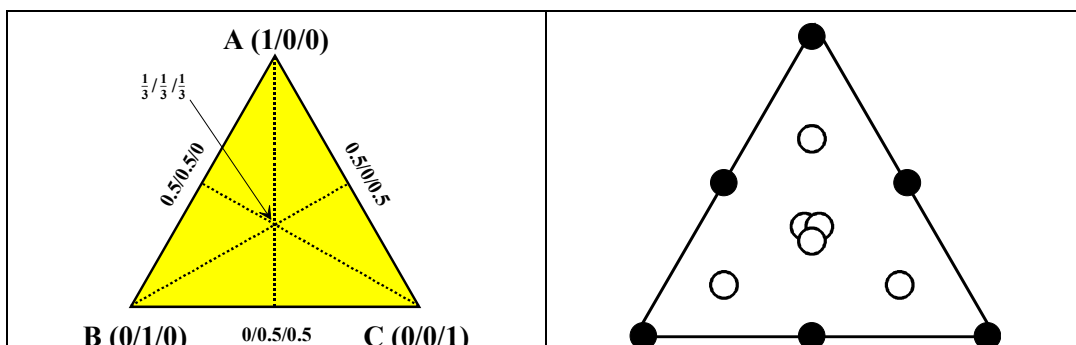
Figure 19.4: (left) Simplex-shaped experimental region.
Figure 19.5: (right) Experimental design used in tablet manufacturing example. Solid points indicate mandatory experiments, open circles optional runs. Note that only one experiment was carried out at the overall centroid.

Step 5: Evaluation of size and shape of mixture region

In the fifth step of the strategy, the size and the shape of the mixture region is considered. This is particularly important when the mixture region is irregular. We will use the tablet example to illustrate this point. One useful approach to checking the size and the shape of the mixture region, and thus to understanding how and where the experiments are planned, might be to (1) generate a preliminary design, (2) fill the experimental worksheet with artificial response data, (3) calculate a (nonsense) regression model, (4) and make contour plots for graphical evaluation of the experimental region.
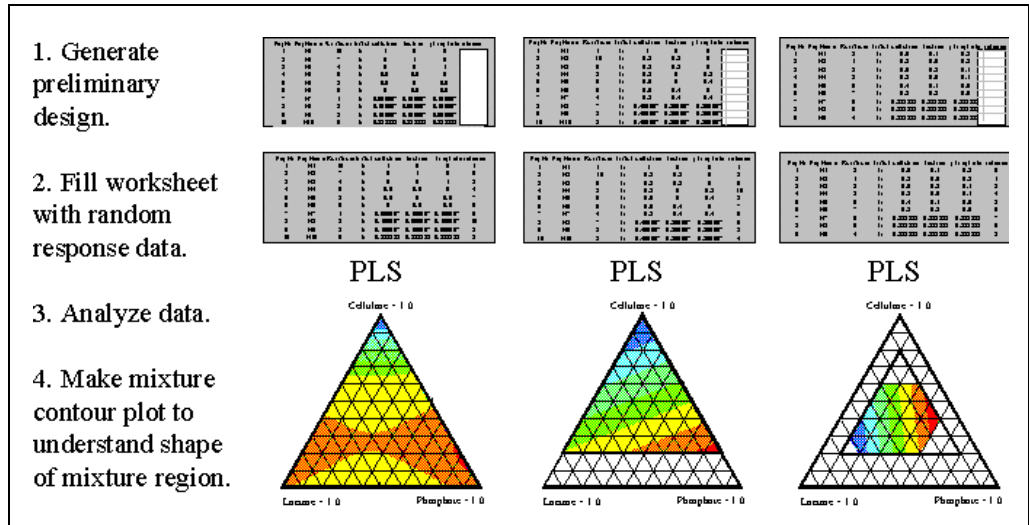
Figure 19.6: *Illustration of the procedure used for understanding the size and the shape of a mixture region.*

The proposed procedure is illustrated in Figure 19.6 for cases based on three mixture factors, a situation resembling the tablet manufacturing application. Three situations, the first with a full-sized simplex-shaped region, the second with a constrained simplex-shaped region, and the third with a constrained irregular region, are depicted. In Figure 19.7 alternative mixture designs for a regular mixture region are shown. As is easily realized, the choice of regression model will be important.
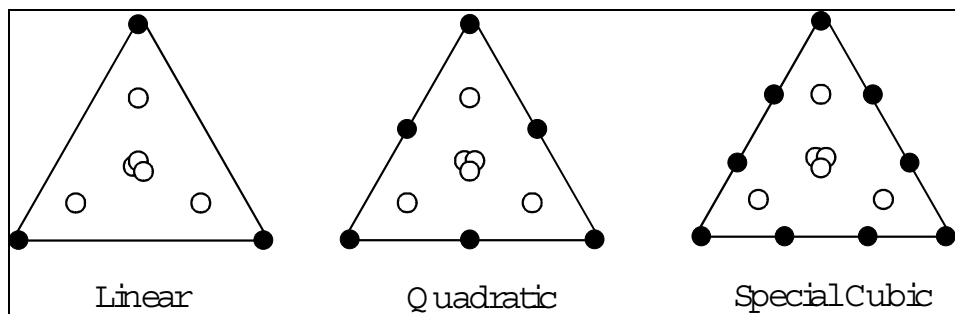
Figure 19.7: *Alternative mixture designs for a regular mixture region in three factors. Open circles denote optional runs.*

Step 6: Definition of reference mixture

The sixth step of the strategy focuses on the definition of the reference mixture. This reference mixture plays an important role in the regression analysis, because the regression coefficients are expressed in relation to its co-ordinates. In a non-constrained simplex-shaped mixture region, the reference mixture corresponds to the overall centroid, that is, the 1/3,1/3,1/3 mixture. This is shown in Figure 19.8. For a regular mixture region, the identification of the reference mixture is facile. However, with irregular experimental regions the task is more taxing, and an efficient algorithmic approach is often needed. The co-ordinates of the reference mixture may well be used for replication. However, the problem is often that the reference mixture is quoted to 4-5 decimal places, which might be difficult to achieve in practice. Thus, it might be necessary to manually modify the calculated reference mixture and round off to a more manageable precision.
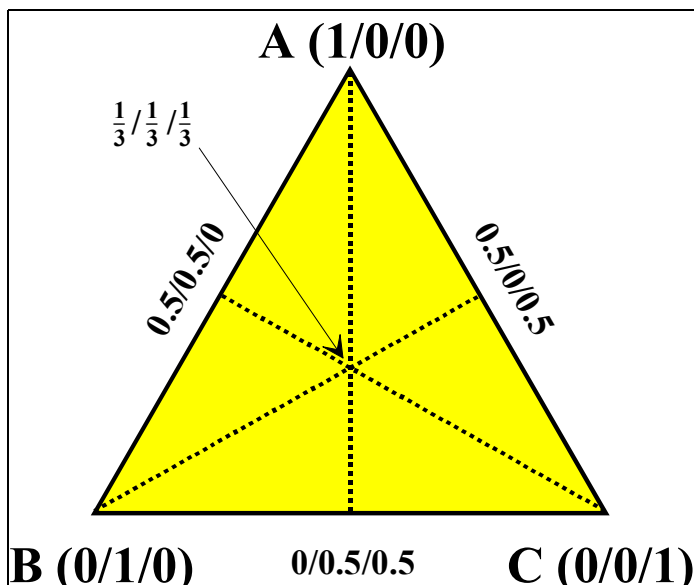
Figure 19.8: *Graphical illustration of the location of the overall centroid in the case of a non-constrained simplex-shaped mixture region.*

Step 7: Execution of design

Once the mixture design has been generated, evaluated, and found to encode a region of sufficient size and acceptable shape, and the reference mixture has been found and possibly adjusted, the next step is to carry out the experiments. It is important that these experiments, as far as is practically possible, are conducted in random order. This is in order to convert any systematic time trend which might occur into random unsystematic behavior. Such unintentionally influencing trends might, for example, be an "improvement" in the laboratory skill of the experimenter, the degrading or aging of an HPLC-column, a constantly decreasing ambient temperature, and so on. Figure 19.9 displays the tablet manufacturing worksheet. It contains a column giving a random run order, which could be employed if these experiments were carried out anew.

Figure 19.9: *The experimental worksheet of the tablet manufacturing application.*

Step 8: Analysis of data and evaluation of model

Step 8 of the strategy is the analysis of data and the evaluation of model. This step is carried out according to the principles outlined in Chapters 1 – 12, the difference being that the PLS regression technique is used instead of MLR. PLS is detailed in Chapter 24 and in the Statistical Appendix. It is appropriate to commence with the evaluation of the raw data. The histogram shown in Figure 19.10 indicates that the response data can be used in the untransformed metric. Furthermore, since the design does not contain any replicated experiments, the replicate plot will not give any information about the replicate error.
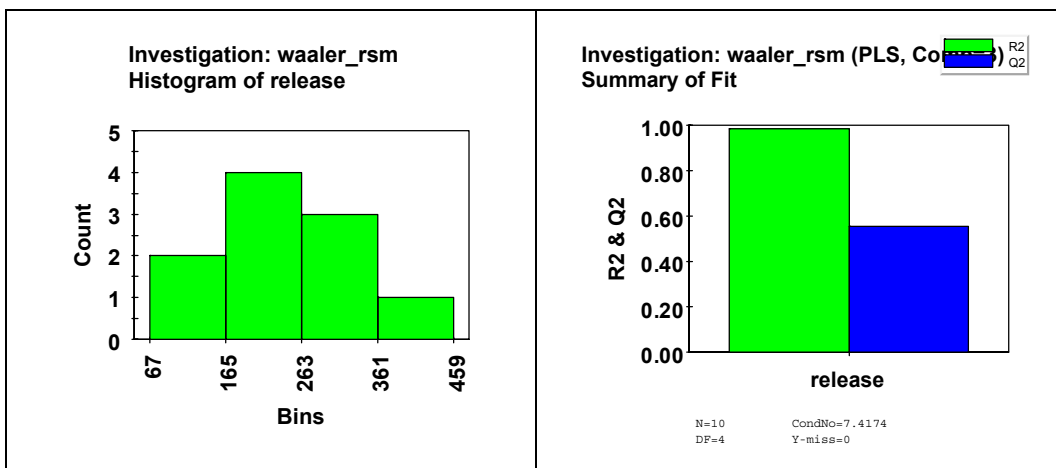
Figure 19.10: *(left) Histogram of Release response.*
Figure 19.11: *(right) Summary of fit of quadratic model fitted to Release.*

The PLS analysis of the tablet data gave a model with $R^2 = 0.98$ and $Q^2 = 0.55$ (Figure 19.11). These statistics point to an imperfect model, because R^2 substantially exceeds Q^2. Unfortunately, the second diagnostic tool (Figure 19.12), the ANOVA table, is incomplete

because the lack of fit test could not be performed. However, a possible reason for the poor modelling is found when looking at the N-plot of the response residuals given in Figure 19.13. Experiment number 10 is an outlier and degrades the predictive ability of the model. If this experiment is omitted and the model refitted, Q^2 will increase from 0.55 to 0.69. We decided *not* to remove the outlier, primarily to conform with the modelling procedure of the original literature source.

release	DF	SS	MS (variance)	F	p	SD
Total	10	624571	62457.102			
Constant	1	532224.875	532224.875			
Total Corrected	9	92346.125	10260.681			101.295
Regression	5	90944.82	18188.965	51.92	9.92E-04	134.866
Residual	4	1401.302	350.325			18.717
Lack of Fit (Model Error)	4	------	------	------	------	------
Pure Error (Replicate Error)	0	------	------			------

N = 10	Q2 = 0.5525	CondNo = 7.4174	
DF = 4	R2 = 0.9848	Y-miss = 0	
Comp= 3	R2Adj = 0.9659	RSD = 18.717	
		ConfLev= 0.95	
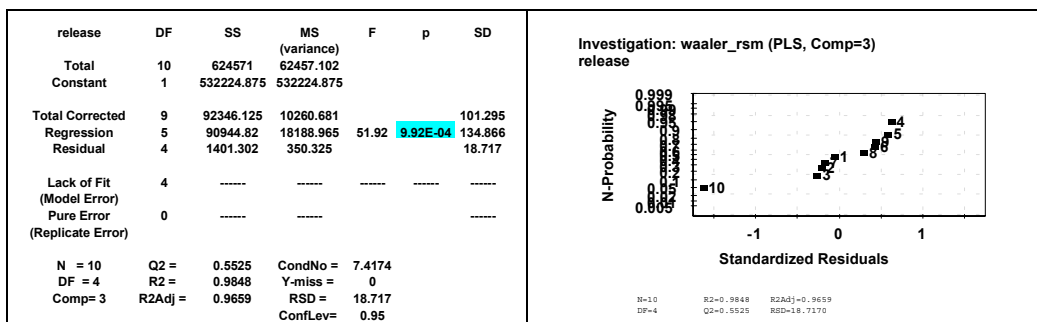
Figure 19.12: *(left) ANOVA of Release.*
Figure 19.13: *(right) N-plot of residuals of Release.*

Step 9: Visualization of modelling results

The ninth step of the strategy concerns the visualization of the modelling results. Scaled and centered regression coefficients of the computed model are plotted in Figure 19.14. This coefficient plot shows that in order to maximize the release rate, the amount of lactose in the recipe should be kept low and the amount of phosphate high. The presence of significant square and interaction terms indicate the existence of quadratic behavior and non-linear blending effects. These effects are more easily understood by means of the trilinear mixture contour plot shown in Figure 19.15. This plot suggests that with the mixture composition 0.32/0/0.68 one may expect a response value above 350. This point should be tested in reality, thus functioning as an experimental verification of the model.
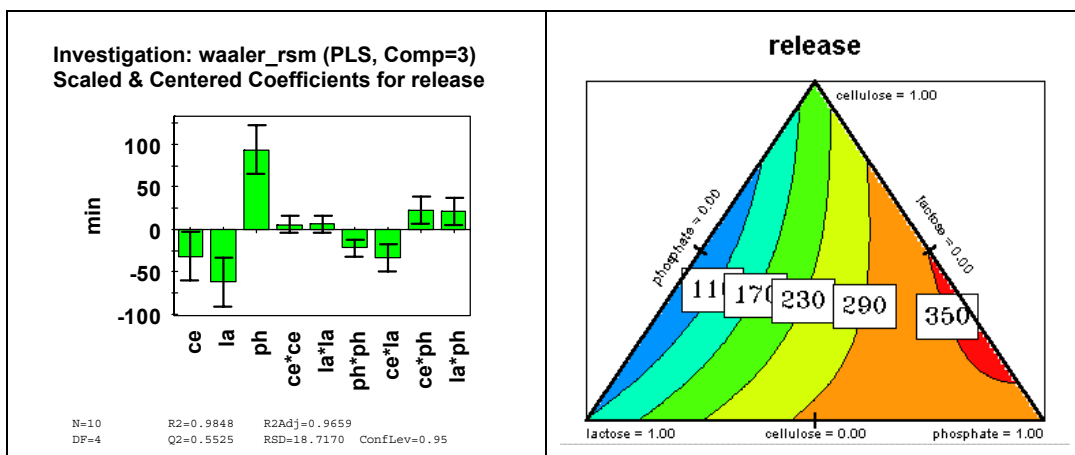
Figure 19.14: *(left) Regression coefficient plot of model for Release.*
Figure 19.15: *(right) Trilinear mixture contour plot for Release.*

Step 10: Use of model

Transferring the modelling results into decisions, plans, and concrete action, is of utmost importance in mixture design. This is the essence of step 10 of the strategy, called use of model. With a screening design one may, for instance, use the resulting model to get an appreciation of where to carry out future experiments, that is, how to modify the factor ranges and thus "move" the experimental domain towards a more interesting region. This region may be mapped with another screening design or an RSM design. In order to accomplish this, the analyst will have to specify a profile of desired response values and then find out which settings of the mixture and/or process factors best correspond to the desired profile. The identification of the optimal factor settings is found with the software optimizer.

In the tablet example, the optimizer identified only one point, the mixture 0.32/0/0.68, where maximum release rate was predicted at 363 minutes. This point was not tested in the original work, but one close to it was. The experimenters performed three verifying experiments and these results together with model predictions are summarized in Figure 19.16. As seen, the model predicts well except for the mixture 0.5/0.125/0.375.

Pred No	cellulose	lactose	phosphate	release (obs)	release(pred)	Lower	Upper
1	0.32	0	0.68	---	363	322	404
2	0.5	0.125	0.375	370	293	262	324
3	0.333	0	0.667	340	363	322	405
4	0.667	0	0.333	345	320	278	361

Figure 19.16: Model predictions of Release.

Advanced mixture designs

Usually, it is not relevant to allow mixture factors to vary between 0 and 1, and other, narrower bounds are therefore employed. Figure 19.17 illustrates a more complex mixture region arising when three mixture factors are constrained. Here, factor A is allowed to vary between 0.2 and 0.6, factor B between 0.1 and 0.6, and factor C between 0.1 and 0.6. This experimental region is irregular and cannot be addressed by any classical mixture design. A D-optimal design is necessary. Another complicated problem which is becoming increasingly addressed is to vary both process factors and mixture factors within the same experimental protocol. An example of this latter case was outlined in Chapter 18.
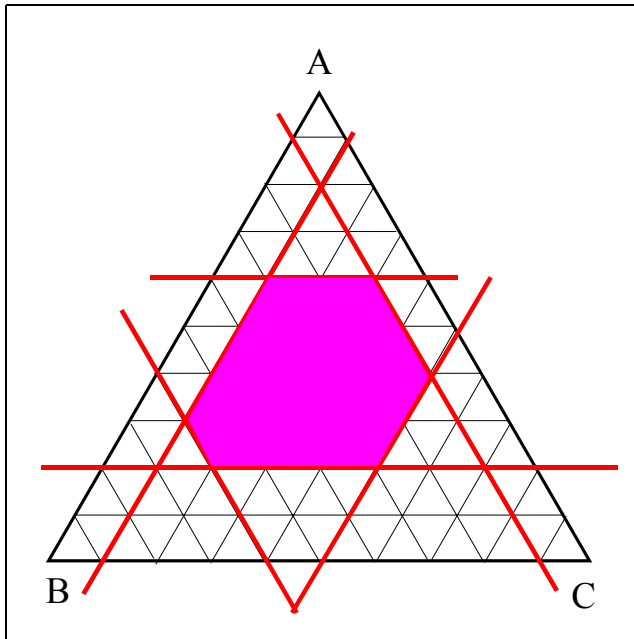
Figure 19.17: *An irregular mixture region in three mixture factors.*

Quiz

Please answer the following questions:

What is a mixture in the context of DOE?

What is a factor bound?

Which experimental objectives can be addressed in mixture design?

What is an extreme vertex?

What is the overall centroid?

What is the reference mixture?

Which three diagnostic tools may be used in the regression modelling?

How can the modelling results be converted into concrete action?

Which kind of design must be used when the experimental region is irregular?

Summary

In this chapter we have provided an introduction to mixture design. The characteristic feature of a mixture is that all ingredients sum to 100%. This imposes a restriction on the factors, so that they can no longer be varied independently of one another. Hence, orthogonal factor arrays no longer apply, and instead mixture designs must be used. We used a tablet manufacturing application to illustrate how mixture design and analysis of data may be carried out in the case when all factors vary between 0 and 1. We also outlined a ten-step strategy for mixture design. By way of example, it was demonstrated that the same

kind of diagnostic and modelling tools, which were introduced for ordinary DOE designs, are useful also for mixture design data. Finally, this chapter ended with a brief look at more complicated, and perhaps more realistic, applications, where the mixture design region is irregular. For such situations there are no classical mixture designs available, and the problems must be addressed with D-optimal designs.

20 Multi-level qualitative factors (Level 3)

Objective

Qualitative factors for which more than two levels have been defined have certain special properties that must be considered in the regression modelling. It is the objective of this chapter to explain how to deal with designs containing such factors. For this purpose, we will concentrate on the use of the regression coefficient plot and the interaction plot. The coefficient plot is a little trickier to interpret than in the case of only quantitative factors or with qualitative factors at two levels. The interaction plot is useful as it facilitates model interpretation. In order to illustrate these modelling principles, we will study a data set related to cotton cultivation, which contains two qualitative factors varied at four and seven levels.

Introduction

Usually, a full factorial design constructed in qualitative factors at many levels is experimentally laborious. Imagine a situation where two qualitative factors and one quantitative factor are varied. Factor A is a qualitative factor with four levels, factor B a qualitative factor with three settings, and factor C a quantitative factor varying between –1 and +1. A full factorial design in this case would correspond to 4*3*2 = 24 experiments, which is depicted by all filled and open circles in Figure 20.1.
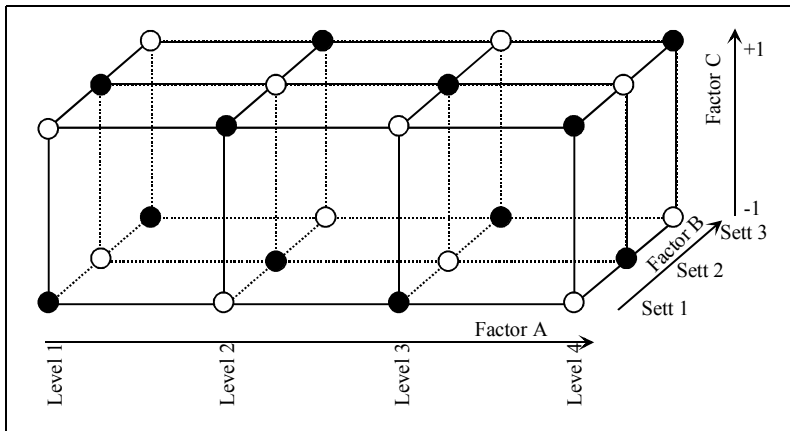
*Figure 20.1: A 4*3*2 full factorial design comprising 24 runs.*

However, with a screening objective and a linear model, the full factorial design displayed in Figure 20.1 requires unnecessarily many runs, and an alternative design with fewer experiments might therefore be legitimate. It is possible to identify such a reduced design with D-optimal techniques (see discussion in Chapter 18 and below).

We will now try to illustrate how one may *analyze* data originating from a design laid out in qualitative factors only.

Example – Cotton cultivation

For a demonstration of data analysis involving multi-level qualitative factors, we will use an example dealing with cotton cultivation. The Cotton application has two factors, which are overviewed in Figure 20.2. One factor is related to the Variety, or species, of cotton, and is varied in four levels, denoted V1 – V4. The other factor represents the Center, or location, of cotton cultivation. For this factor, as many as seven levels are defined, termed C1 – C7.

Factors

	Name	Abbr.	Units	Type	Use	Settings
1	**Variety**	V		Qualitative	Controlled	V1,V2,V3,V4
2	**Center**	C		Qualitative	Controlled	C1,C2,C3,C4,C5,C6,C7

Responses

	Name	Abbr.	Units	Transform
1	**Yield**	Yi		None

Figure 20.2: *(left) Overview of factors varied in the Cotton example.*
Figure 20.3: *(right) The Yield response measured in the Cotton example.*

To describe the growth of cotton a response called Yield was measured. Figure 20.3 summarizes this response. The Yield was registered relative to a standard crop, and a high numerical value designates good growth. To investigate the impact of the two factors on the Yield, the experimenters conducted a full factorial design in 28 runs. The worksheet for this design is given in Figure 20.4. Because of the abundance of experimental data points it is possible to fit an interaction model.

Worksheet

	1	2	3	4	5	6	7
	Exp No	Exp Name	Run Order	Incl/Excl	Variety	Center	Yield
1	1	N1	23	Incl	V1	C1	-30
2	2	N2	24	Incl	V2	C1	-39
3	3	N3	16	Incl	V3	C1	-34
4	4	N4	11	Incl	V4	C1	-24
5	5	N5	27	Incl	V1	C2	4
6	6	N6	15	Incl	V2	C2	5
7	7	N7	13	Incl	V3	C2	-4
8	8	N8	1	Incl	V4	C2	10
9	9	N9	3	Incl	V1	C3	-20
10	10	N10	5	Incl	V2	C3	-21
11	11	N11	9	Incl	V3	C3	-25
12	12	N12	28	Incl	V4	C3	-13
13	13	N13	6	Incl	V1	C4	53
14	14	N14	26	Incl	V2	C4	47
15	15	N15	20	Incl	V3	C4	49
16	16	N16	7	Incl	V4	C4	76
17	17	N17	4	Incl	V1	C5	-33
18	18	N18	19	Incl	V2	C5	-31
19	19	N19	8	Incl	V3	C5	-34
20	20	N20	10	Incl	V4	C5	-26
21	21	N21	21	Incl	V1	C6	8
22	22	N22	14	Incl	V2	C6	6
23	23	N23	17	Incl	V3	C6	0
24	24	N24	2	Incl	V4	C6	14
25	25	N25	18	Incl	V1	C7	16
26	26	N26	12	Incl	V2	C7	12
27	27	N27	22	Incl	V3	C7	11
28	28	N28	25	Incl	V4	C7	16

Figure 20.4: *Experimental data of the Cotton application.*

Regression analysis of Cotton application – Coefficient plot

We will now relate what happened when fitting an interaction model to the Yield response. It is emphasized that our primary aim is to highlight certain technical characteristics of

qualitative factors, and not so much to provide a detailed account of the regression analysis. The interaction model in question will have four parent terms, the constant term, the two main effects of Variety and Center, and their two-factor interaction. When fitting this interaction model to the data, the regression coefficient plot shown in Figure 20.5 was acquired. This plot has 39 bars, each representing one coefficient. Four bars originate from Variety, seven bars from Center, and 28 from their two-factor interaction. Note that we do not give a summary of fit plot because the model is saturated and hence R^2 is 1.0.
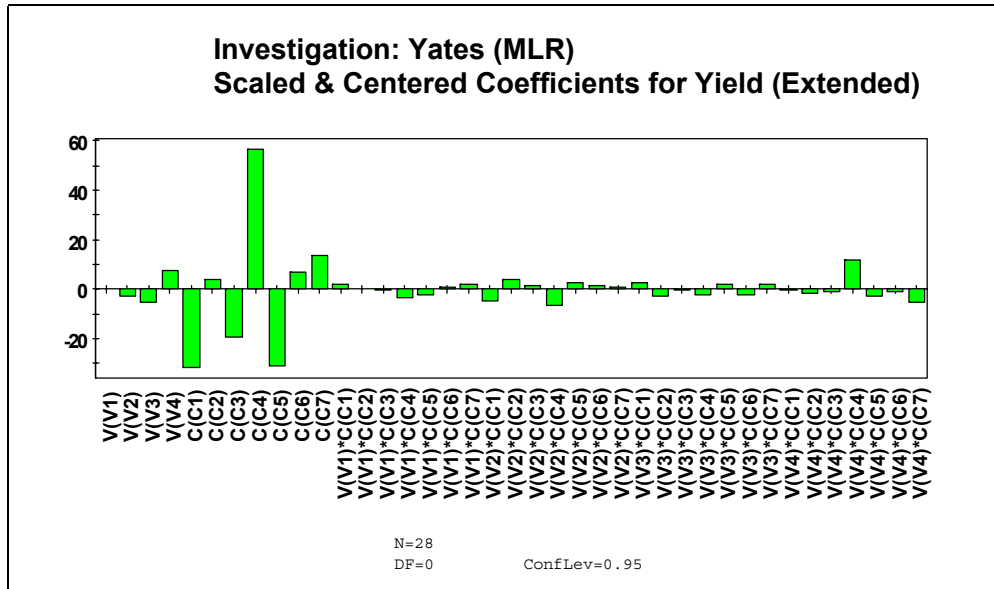
Figure 20.5: Regression coefficient plot of the interaction model used in the Cotton example.

The abundance of regression coefficients is a characteristic feature of designs containing multi-level qualitative factors. Furthermore, each block of regression coefficients, that is, those arising from Center, Variety, and so on, must be treated as a single entity. For instance, it is not possible to delete only some of the 28 coefficients pertaining to the V*C two-factor interaction. On the contrary, these terms must be included in, or removed from, the model as an aggregate.

We will now interpret the model. Although the coefficient plot is cluttered, its message is that Center has more impact on the result than Variety. A particularly good Yield is achieved with Center #4.

Regression analysis of Cotton application – Interaction plot

In the case of many multi-level qualitative factors, the regression coefficient plot tends to be cluttered and hard to overview. Fortunately, there is another graphical tool available which will facilitate model interpretation. This is the so called interaction plot. Figure 20.6 displays such an interaction plot of the model fitted to the cotton cultivation data. With this graph it does not take long to realize that the best possible combination of factors is Variety #4 and Center #4. In addition, it appears that change of cultivation center has more impact on the result than change of cotton variety.
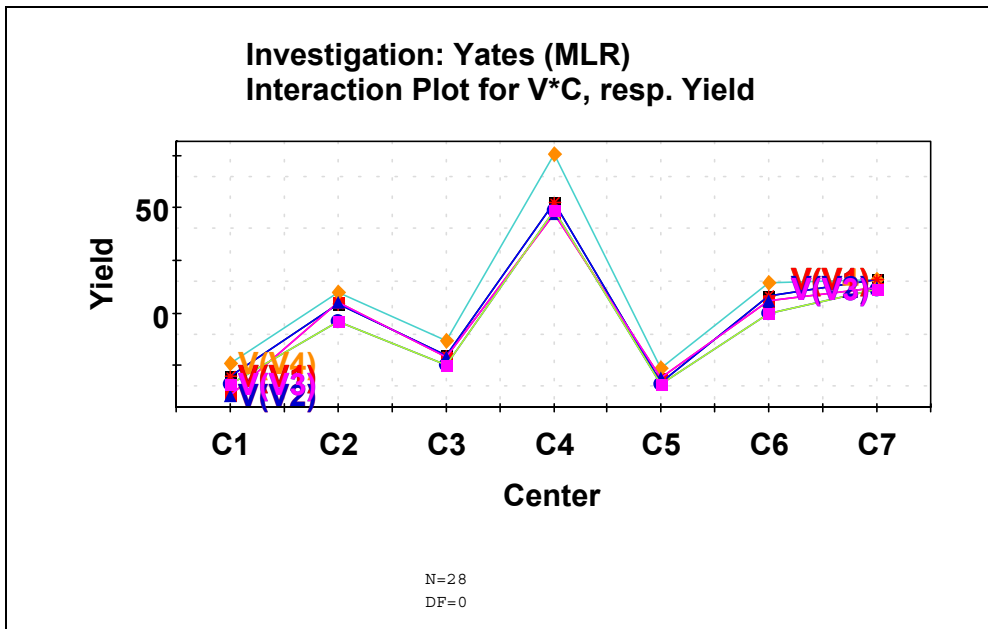
Figure 20.6: Interaction plot of the interaction model used in the Cotton example.

This concludes our presentation of the modelling results. In the following, our intention is to describe the reason why multi-level qualitative factors appear as they do in a coefficient plot.

Regression coding of qualitative variables

Qualitative variables require a special form of coding for regression analysis to work properly. More specifically, a qualitative factor with k levels, will have k-1 expanded terms in the model calculations. Let us consider the Variety factor in the Cotton application. This factor has four levels and requires three degrees of freedom, three experiments, to be estimable. In order to be able to estimate its impact a slight mathematical re-expression is necessary prior to the data analysis. This pre-treatment is outlined in Figure 20.7.

	Expanded term		
Level of factor	V(V2)	V(V3)	V(V4)
V1	-1	-1	-1
V2	1	0	0
V3	0	1	0
V4	0	0	1

Figure 20.7: Regression coding of the four-level qualitative factor Variety.

We can see that the Variety factor is expanded into three artificial categorical variables. In a similar manner, the Center factor of seven levels, will demand six categorical variables for coding its information. In regression modelling, each such expanded term gives rise to one regression coefficient, and may therefore be visually inspected in a regression coefficient plot.

Regular and extended lists of coefficients

All expanded model terms related to a qualitative factor are presented either according to the principle of *regular coefficients* display or to the principle of *extended coefficients* display. This is exemplified in Figure 20.8. In the *regular* mode, the coefficients of the expanded terms of Variety are given as the coefficients for level 2, V2, level 3, V3, and level 4, V4. In the *extended* mode the coefficient for level 1, V1, is computed as the negative sum of the coefficients of the other expanded terms. This summation is shown in Figure 20.8. Because the four coefficients sum to zero, only three degrees of freedom are needed for Variety.

Regular			Extended	
Yield	**Coeff.**		**Yield**	**Coeff.**
Constant	**-0.25**		**Constant**	**-0.25**
			V	**DF = 3**
			V(V1)	**-0.035**
V(V2)	**-2.75**		**V(V2)**	**-2.75**
V(V3)	**-5.036**		**V(V3)**	**-5.036**
V(V4)	**7.821**		**V(V4)**	**7.821**
Sum	*0.035*		*Sum*	*0*
			C	**DF = 6**
			C(C1)	**-31.5**
C(C2)	**4**		**C(C2)**	**4**
C(C3)	**-19.5**		**C(C3)**	**-19.5**
C(C4)	**56.5**		**C(C4)**	**56.5**
C(C5)	**-30.75**		**C(C5)**	**-30.75**
C(C6)	**7.25**		**C(C6)**	**7.25**
C(C7)	**14**		**C(C7)**	**14**
Sum	*31.5*		*Sum*	*0*

Figure 20.8: Illustration of regular and extended lists of regression coefficients.

In the lower part of Figure 20.8, an analogous presentation of the coefficients related to the expansion of Center is given. We can see that six degrees of freedom are sufficient for the seven-level factor. The V*C two-factor interaction is expanded and computed in a similar manner. This means that 3*6 =18 or 4*7 = 28 coefficients would be displayed, depending on whether regular or extended mode is used. In the analysis, it is up to the user to specify whether regular or extended mode is preferable. When the extended mode is used, a total of 39 coefficients are displayed for the Cotton application, namely 4 for variety, 7 for Center, and 28 for their interaction (Figure 20.5).

Generation of designs with multi-level qualitative factors

The last topic that we will address in this chapter concerns design generation. With only two-level qualitative factors the design generation is simple, since regular full and fractional factorial designs are readily available. When a qualitative factor is explored at three or more levels, however, the task of creating a good experimental design with few runs becomes more problematic.

Let us go back to the example which we described at the beginning of this chapter, which had one four-level and one three-level qualitative factor, plus a quantitative factor. A full factorial design for screening would in this case correspond to 4*3*2 = 24 experiments, which is depicted by all filled and open circles in Figure 20.9. In addition to these mandatory "factorial" experiments, between 3 and 5 replicates ought to be appended to the final design. Hence, between 27 and 29 experiments would be necessary, which are too many for a screening objective. An alternative design with fewer experiments would be preferred. Such an alternative design in only 12 experiments is given by the solid circles, or equivalently, by the set of unfilled circles. Observe that replicated experiments are not included in the selected dozen, but must be added.
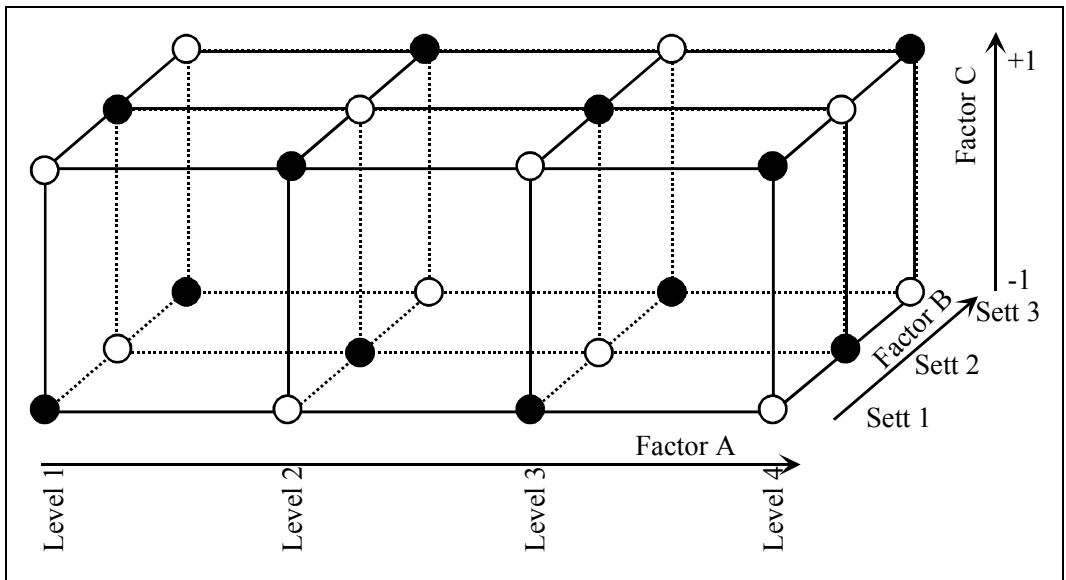
Figure 20.9: *A 4*3*2 full factorial design comprising 24 runs (filled* and *open points), or a related reduced design in 12 runs (filled* or *open points).*

These subset selections shown in Figure 20.9 were made using a theoretical algorithm, a D-optimal algorithm, for finding those experiments with the best spread and best balanced distribution. A balanced design has the same number of runs for each level of a qualitative factor. This feature is not absolutely necessary, but often convenient and makes the design easier to understand. D-optimal designs were discussed in Chapter 18.

Quiz

Please answer the following question:

What is a multi-level qualitative factor?

Which type of plot is particularly informative for evaluating multi-level qualitative factors?

How many degrees of freedom are associated with the main term of a five-level qualitative factor?

How many coefficients of expanded terms of a five-level qualitative factor are encountered in regular coefficient mode? In extended mode?

Which type of design is useful for multi-level qualitative factors?

Summary

In this chapter, we have described some interesting features pertaining to experimental design with multi-level qualitative factors. One characteristic property of multi-level qualitative factors is that their expanded terms are linked to one another. Hence, each such block of associated terms must be treated as an aggregate, which cannot be split up to handle individual terms separately. In connection with this, we also demonstrated that regression coefficient plots may be presented in two styles, in regular mode or extended mode. Another feature of qualitative factors is that full factorial designs are usually not useful, because they require too many experimental trials. Rather, D-optimal designs are more appropriate.

21 Taguchi approach to robust design (Level 3)

Objective

The objective of this chapter is to describe the Taguchi approach to robust design. "Robust design" is a term with a confusing similarity to "robustness testing", discussed in Chapter 17. In the Taguchi approach, robustness has a different connotation and objective. The objective is to find conditions where simultaneously the responses have values close to target and low variability. Hence, factors are often varied in large intervals and with designs very different from those discussed in Chapter 17. This contrasts with robustness testing where small factor intervals are used. The Taguchi philosophy aims to reduce variation in the process outcomes by finding operating conditions under which uncontrollable variation in some process factors has minimal impact on the product quality. It is the aim of this chapter to convey the essence of this approach. Taguchi distinguishes between *design factors* which are easy-to-control factors, and *noise factors* which are hard-to-control factors. These factors are usually varied in so called inner and outer arrays. We will use the full CakeMix application, and an additional example from the pharmaceutical industry to illustrate the steps of the Taguchi approach. Two ways to analyze Taguchi arrayed data, the classical analysis approach and the interaction analysis approach, are considered.

The Taguchi approach – Introduction

The original ideas of robust design were formulated by the Japanese engineer Genichi Taguchi. The Taguchi approach made rapid progress in Japanese industry, and partly explains the success of the Japanese electronics and car industries in the 1970s and 1980s. Taguchi divides robust design into three major blocks, the *product* design, the *parameter* design, and the *tolerance* design phases.

Product design involves the use of off-line quality improvement schemes. Taguchi advocates a philosophy where quality is measured in terms of loss suffered by society as a result of product variability around a specified target. This includes losses to the manufacturer during production, and losses to the consumer after the release of the product. Every contribution is quantified in monetary terms and the total loss is computed. A desirable product is one for which the total loss is acceptably small. This means that high quality is coupled to low loss. In this context, Taguchi recommends the use of a loss function for estimating the loss. The problem is that the computation of such a loss expression is often obscured by hurdles in the definition of relevant target values and an appropriate form of the polynomial model. Often wide-ranging assessments of the product

impact on society must be made. If successful, this may lead to the specification of an acceptable region within which the final design can lie. It is in this region that the second part of the Taguchi approach, the parameter design, is applied.

The parameter design is equivalent to using design of experiments (DOE) for finding optimal settings of the process variables. The additional feature of the Taguchi approach is to make a distinction between *design* factors and *noise* factors. The former are easy-to-control and are supposed to affect the mean output of the process. The latter are hard-to-control and may or may not affect the process mean, and the spread around the mean. The subsequent data analysis tries to identify design factors which affect only the mean, design factors which affect only the spread around the mean, and factors which affect both properties. The idea is then to choose levels of the design factors so that the process output is high and relatively insensitive to the noise factors. However, one limitation of the Taguchi approach concerning this second stage, is that the prescribed strategy is weak on the modelling-of-data aspect.

Finally, the third part of the Taguchi approach, the tolerance design, takes place when optimal factor settings have been specified. Should the variability in the product quality still be unacceptably high, the tolerances on the factors have to be further adjusted. Usually, this is accomplished by using a derived mathematical model of the process, and the loss function belonging to the product property of interest.

Arranging factors in inner and outer arrays

In the Taguchi approach, *design factors* and *noise factors* are varied systematically according to an architecture consisting of inner and outer factor arrays. We will now study this arrangement. For this purpose, we will use an extended version of the CakeMix application. We recall that this is an industrial pilot plant investigation aimed at designing a cake mix giving tasty products. Previously, we studied how changes of the factors Flour, Shortening, and Eggpowder affected the taste. In reality, this example is nothing less than a robust design application, where the final goal is to design a cake mix which will produce a good cake even if the customer does not follow the baking instructions. To explore whether this was feasible, the factors Flour, Shortening, and Eggpowder were used as design factors and varied in a cubic inner array (see Figure 21.1). In addition, two noise factors were incorporated in the experimental design as a square outer array. These factors were baking temperature, varied between 175 and 225°C, and time spent in oven, varied between 30 and 50 minutes.
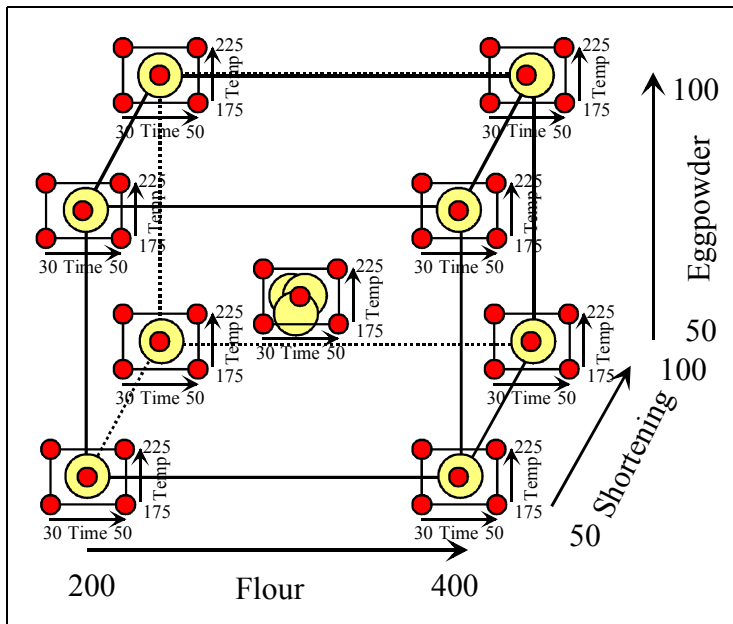
Figure 21.1: The arrangement of the CakeMix factors as inner and outer arrays.

We can see in Figure 21.1 that for each design point in the inner cubic array, a square outer array is laid out. The number of experiments according to the inner array is 11, and according to the outer array 5, which makes 11*5 = 55 experiments necessary. With this arrangement of the experiments, the experimental goal was to find levels of the three ingredients producing a good cake (a) when the noise factors temperature and time were correctly set according to the instructions on the box, and (b) when deviations from these specifications occur. Hence, in this kind of testing, the producer has to consider worst-case scenarios corresponding to what the consumer might do with the product, and let these considerations regulate the low and high levels of the noise factors. Finally, we have another important aspect to observe in this context – the fact that the noise factors are often controllable in the experiment, but not in full-scale production.

The classical analysis approach

The classical way to analyze Taguchi-like DOE data is to form, for each experimental point in the inner array, two responses. The first response is the average response value and the second the standard deviation around this average. These experimental values are shown in Figure 21.2. The Taste response is in the same shape as before. Thus, this is the average taste for the five outer array experiments at each inner array point. Similarly, the second response, denoted StDev, is the standard deviation among the five outer array experiments at each inner array point. Because such standard deviation responses tend to be non-normally distributed, it is common practice to model the log-transformed variable. Thus, in the following, the StDev response will be log-transformed. With the classical analysis approach the goal is to find which design factors affect the variation (StDev) only, which affect the mean level (Taste) only, and which affect both. Note that with this approach, there will be no model terms related to the noise factors (Time and Temperature). We will now study the results of the regression modelling.

Figure 21.2: *Data arrangement for the classical analysis approach.*

Regression analysis

It is instructive to first consider the raw experimental data. Figures 21.3 and 21.4 show the replicate plots of the responses. We can see that for both responses the replicate error is satisfactorily small. It is also of interest that the responses are inversely correlated, as is evidenced by Figure 21.5. We recall that the experimental goal is a factor combination producing a tasty cake with low variation. Hence, it seems as if experiment number 6 is the most promising.
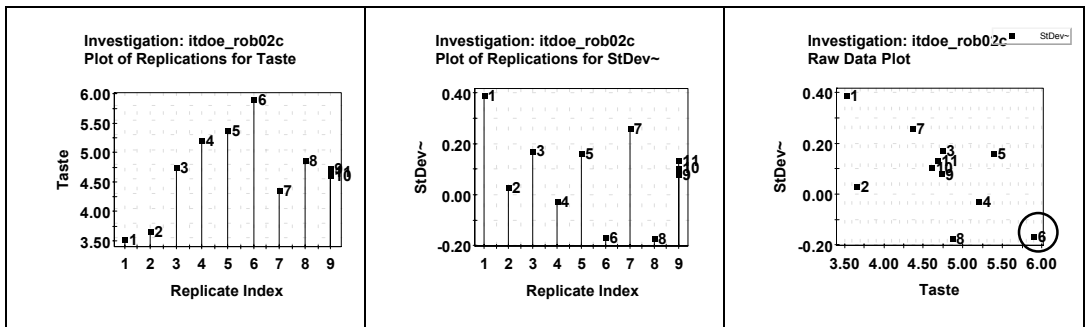
Figure 21.3: *(left) Replicate plot of Taste.*
Figure 21.4: *(middle) Replicate plot of StDev.*
Figure 21.5: *(right) Raw data scatter plot of StDev versus Taste.*

Figures 21.6 – 21.8 show the modelling results obtained from fitting an interaction model to each response. Noteworthy is the negative Q^2 of StDev, indicating model problems. The model for Taste is better, but we remember from previous modelling attempts that even better results are possible if the two non-significant two-factor interactions are omitted.
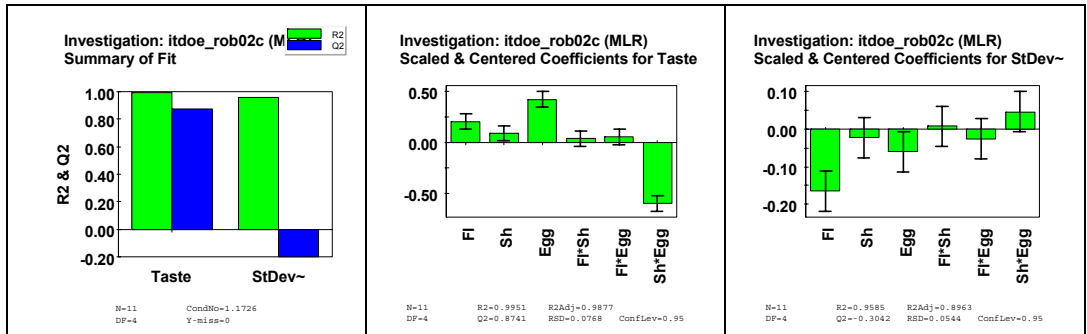
Figure 21.6: *(left) Summary of fit plot of interaction model.*
Figure 21.7: *(middle) Regression coefficient plot for Taste.*
Figure 21.8: *(right) Regression coefficient plot for StDev.*

The results from the fitting of a refined model to each response are seen in Figures 21.9 – 21.11. Obviously, the model for StDev has improved a lot thanks to the model pruning. Two interesting observations can now be made. The first is related to the Sh*Egg interaction, which is much smaller for StDev than for Taste. The second observation concerns the Fl main effect, which shows that Flour is the factor causing most spread around the average Taste. Hence, this is a factor which must be better controlled if robustness is to be achieved. The models that we have derived will now be used to try to achieve the experimental goal.
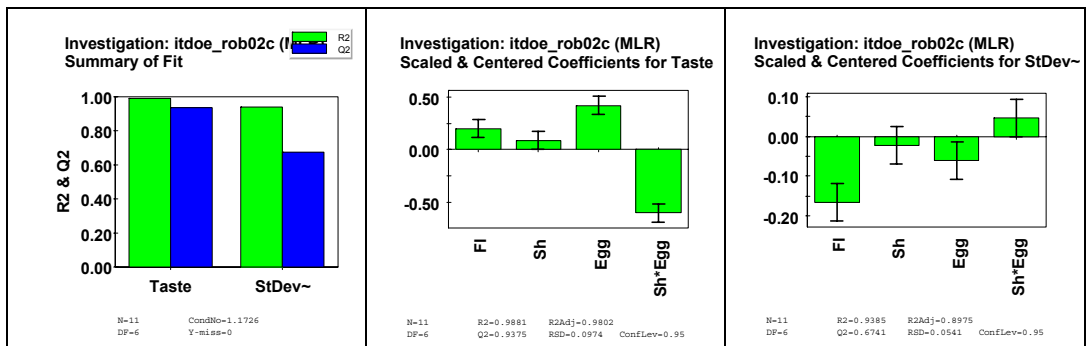
Figure 21.9: *(left) Summary of fit plot of refined interaction model.*
Figure 21.10: *(middle) Regression coefficient plot for Taste – refined model.*
Figure 21.11: *(right) Regression coefficient plot for StDev – refined model.*

Interpretation of model

One way to understand the impact of the surviving two-factor interaction is to make interaction plots of the type shown in Figures 21.12 and 21.13. Evidently, the impact of this model term is greater for Taste than for StDev. This is inferred from the fact that the two lines cross each other in the plot related to Taste, but do not cross in the other interaction plot. Both plots indicate that a low level of Shortening and a high level of Eggpowder are favorable for high Taste and low StDev.
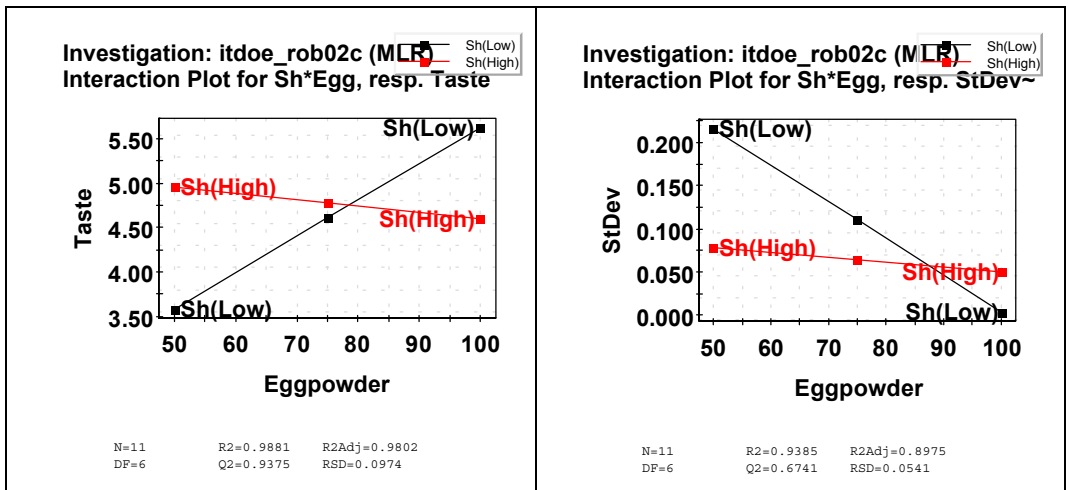
Figure 21.12: *(left) Interaction plot of Sh*Egg with respect to Taste.*
Figure 21.13: *(right) Interaction plot of Sh*Egg with respect to StDev.*

An alternative procedure for understanding the modelled system consists of making response contour plots. Such response contour plots are shown in Figure 21.14. These contours were created by setting Flour to its high level, as this was found favorable in the modelling. The two contour plots convey an unambiguous message. The best cake mix conditions are found in the upper left-hand corner, where the highest taste is predicted to be found, and at the same time the lowest standard deviation. This location corresponds to the factor settings Flour = 400, Shortening = 50, and Eggpowder = 100. At this factor combination, Taste is predicted at 5.84 ± 0.18, and StDev at 0.69 (0.55 - 0.87). Bearing in mind that the highest registered experimental value of Taste is 6.9, and the lowest value of StDev 0.67, these predictions appear reasonable.
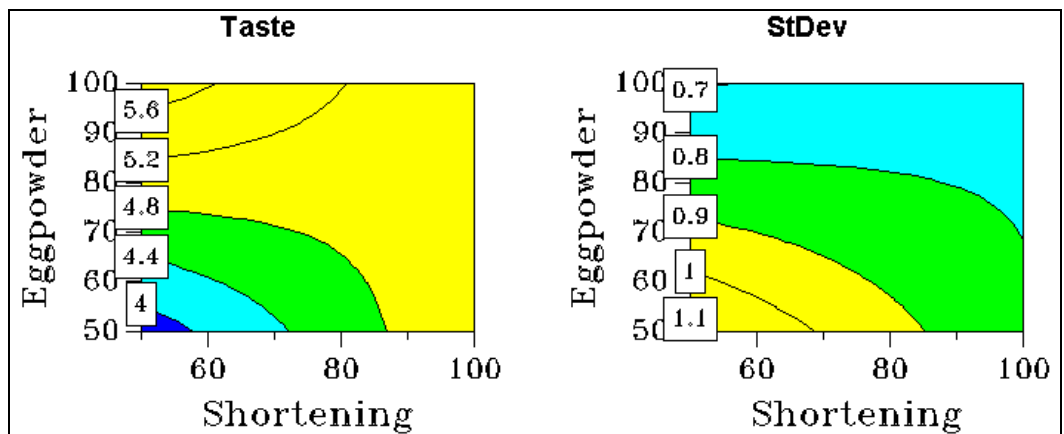
Figure 21.14: *Response contour plots of Taste and StDev (Flour = 400g).*

The interaction analysis approach

One drawback of the classical approach to data analysis is that it does not allow the user to identify which noise factors may be affecting the variability of the responses. For the Taguchi method to be really successful, we would need to be able to estimate the impact of the noise factors *and* possible interactions between the design and the noise factors. Clearly, by definition, the success of the Taguchi approach critically depends on the existence of such noise-design factor interactions. Otherwise, the noise (variability) cannot be reduced by changing any design factors.

This kind of noise/design factor interaction information can be extracted if the design factors and the noise factors are combined into one single design, and a regression model is fitted which contains both types of factors, as well as their interactions. In such a case, what in the classical analysis approach were design factor effects on the variation around the mean, would in this alternative *interaction analysis* approach correspond to noise-design factor cross-terms. We will now use the full CakeMix application to illustrate the interaction analysis approach. For this purpose, we need to rearrange the experimental worksheet, so that it comprises 55 rows. An excerpt of this worksheet is shown in Figure 21.15. As seen, there is only one response, the raw value of Taste. One additional benefit obtained by ordering the experiments in this latter way, is that it enables single deviating experiments to be identified.

| No	Flour	Shortening	Eggpowder	Temp	Time	Taste	No	Flour	Shortening	Eggpowder	Temp	Time	Taste
1	200	50	50	175	30	1.1	34	200	50	50	225	50	1.3
2	400	50	50	175	30	3.8	35	400	50	50	225	50	2.1
3	200	100	50	175	30	3.7	36	200	100	50	225	50	2.9
4	400	100	50	175	30	4.5	37	400	100	50	225	50	5.2
5	200	50	100	175	30	4.2	38	200	50	100	225	50	3.5
6	400	50	100	175	30	5	39	400	50	100	225	50	5.7
7	200	100	100	175	30	3.1	40	200	100	100	225	50	3
8	400	100	100	175	30	3.9	41	400	100	100	225	50	5.4
9	300	75	75	175	30	3.5	42	300	75	75	225	50	4.1
10	300	75	75	175	30	3.4	43	300	75	75	225	50	3.8
11	300	75	75	175	30	3.4	44	300	75	75	225	50	3.8
12	200	50	50	225	30	5.7	45	200	50	50	200	40	3.1
13	400	50	50	225	30	4.9	46	400	50	50	200	40	3.2
14	200	100	50	225	30	5.1	47	200	100	50	200	40	5.3
15	400	100	50	225	30	6.4	48	400	100	50	200	40	4.1
16	200	50	100	225	30	6.8	49	200	50	100	200	40	5.9
17	400	50	100	225	30	6	50	400	50	100	200	40	6.9
18	200	100	100	225	30	6.3	51	200	100	100	200	40	3
19	400	100	100	225	30	5.5	52	400	100	100	200	40	4.5
20	300	75	75	225	30	5.15	53	300	75	75	200	40	6.6
21	300	75	75	225	30	5.3	54	300	75	75	200	40	6.5
22	300	75	75	225	30	5.4	55	300	75	75	200	40	6.7
23	200	50	50	175	50	6.4							
24	400	50	50	175	50	4.3							
25	200	100	50	175	50	6.7							
26	400	100	50	175	50	5.8							
27	200	50	100	175	50	6.5							
28	400	50	100	175	50	5.9							
29	200	100	100	175	50	6.4							
30	400	100	100	175	50	5							
31	300	75	75	175	50	4.3							
32	300	75	75	175	50	4.05							
33	300	75	75	175	50	4.1							

Figure 21.15: Data arrangement for the interaction analysis approach.

Regression analysis

As usual, we commence the data analysis by evaluating the raw data. Figure 21.16 suggests that the replicate error is small, and Figure 21.17 that the response is approximately normally distributed. Hence, we may proceed to the regression analysis phase, without further preprocessing of the data.
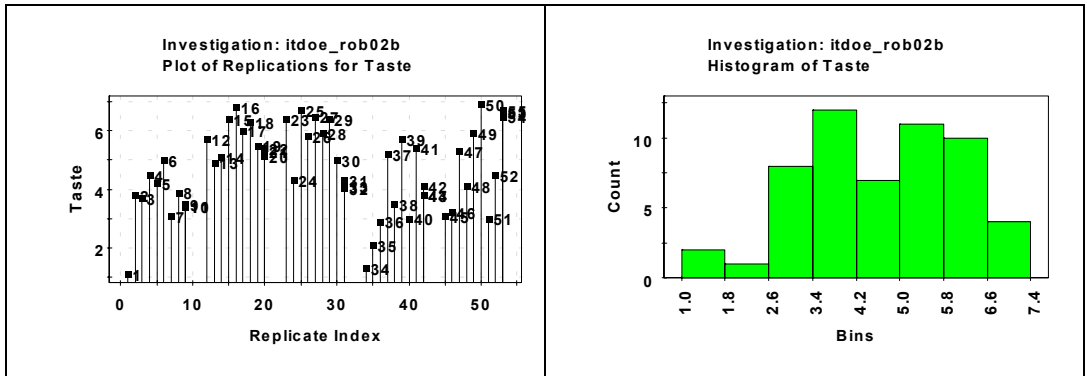

Figure 21.16: *(left) Replicate plot of Taste.*
Figure 21.17: *(right) Histogram of Taste.*

As seen in Figure 21.18, the regression analysis gives a poor model with $R^2 = 0.60$ and $Q^2 = 0.18$. Such a large gap between R^2 and Q^2 is undesirable and indicates model inadequacy. The N-plot of residuals in Figure 21.19 reveals no clues which could explain the poor modelling performance. However, the regression coefficient plot in Figure 21.20 does reveal two plausible causes. Firstly, the model contains many irrelevant two-factor interactions. Secondly, it is surprising that the Fl*Te and Fl*Ti two-factor interactions are so weak. Since in the previous analysis we observed the strong impact of Flour on StDev, we would now expect much stronger noise-design factor interactions. In principle, this means that there must be a crucial higher-order term missing in the model, the Fl*Te*Ti three-factor interaction. Consequently, in the model revision, we decided to add this three-factor interaction and remove six unnecessary two-factor interactions.
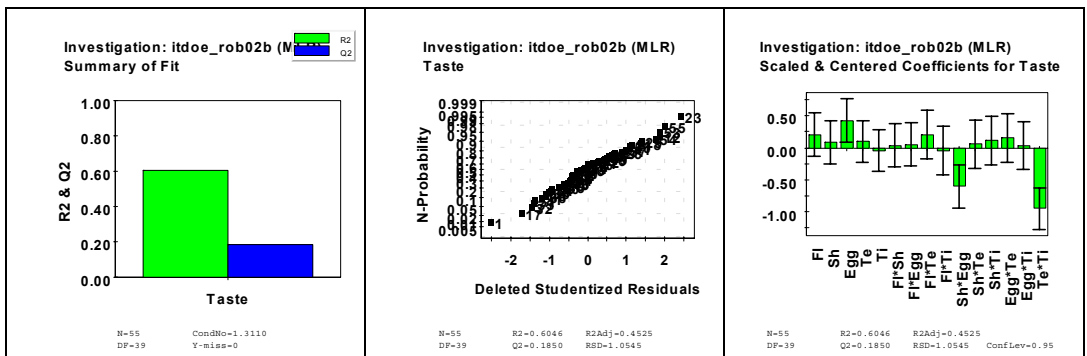

Figure 21.18: *(left) Summary of fit of Taste.*
Figure 21.19: *(middle) N-plot of residuals for Taste.*
Figure 21.20: *(right) Regression coefficients for Taste.*

When re-analyzing the data, a more stable model with the reasonable statistics $R^2 = 0.69$ and $Q^2 = 0.57$ was the result (Figure 21.21). An interesting aspect is that the R^2 obtained is lower than in the classical analysis approach. This is due to the stabilizing effect of averaging Taste over five trials in the classical analysis approach. Concerning the current model, we are unable to detect significant outliers. The relevant N-plot of residuals is displayed in Figure 21.22. Therefore, it is appropriate to consider the regression coefficients, which are displayed in Figure 21.23. We can see the significance of the included three-factor interaction. This is in line with the previous finding regarding the impact of Flour on StDev. Some smaller two-factor interactions are also kept in the model to make the three-factor interaction more interpretable.
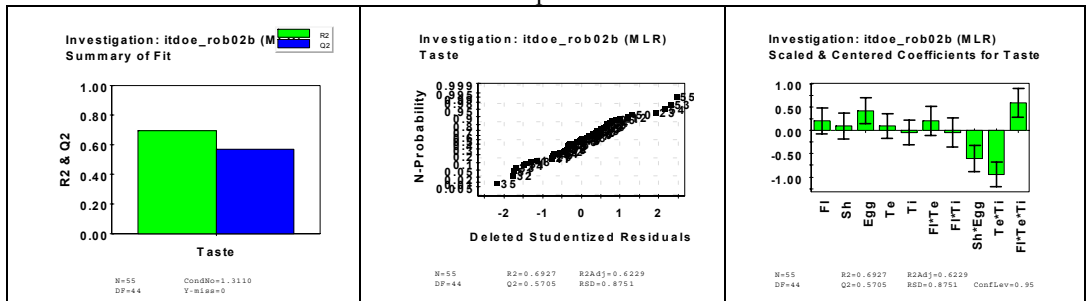
Figure 21.21: (left) Summary of fit of Taste – refined model.
Figure 21.22: (middle) N-plot of residuals for Taste – refined model.
Figure 21.23: (right) Regression coefficients for Taste – refined model.

An important three-factor interaction

The meaning of the three-factor interaction is most easily understood by constructing an interaction plot. Figure 21.24 displays the impact of the three-factor interaction. Now, one may wonder, what is it that we should look for in this kind of plot? The answer is that we want to understand how to adjust the controllable factor Flour, so that the impact of variations in the uncontrollable factors Temperature and Time is minimized. It is seen in Figure 21.24 that by adjusting Flour to 400g the spread in Taste due to variations in Temperature and Time is limited.

Figure 21.24: Interaction plot showing the significance of the Fl*Te*Ti three-factor interaction.

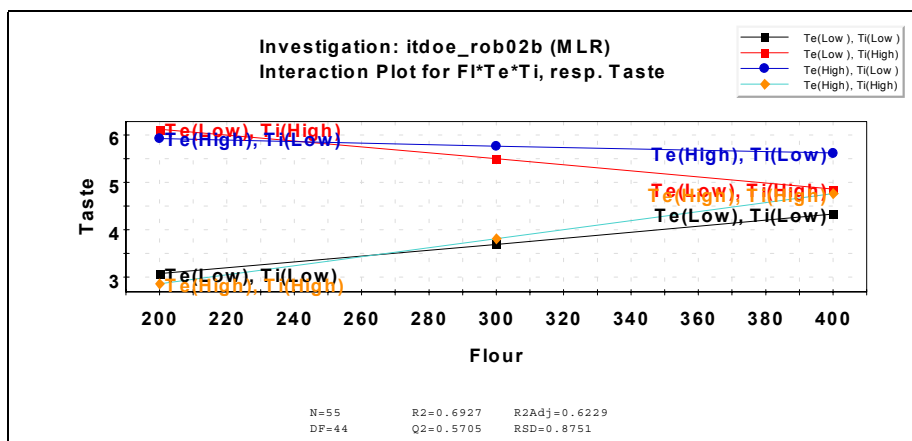Furthermore, in solving the problem, we must not forget the significance of the strong Sh*Egg two-factor interaction. We know from the initial analysis that the combination of low Shortening and high Eggpowder produces the best cakes. These considerations lead to the construction of the response contour triplet shown in Figure 21.25. Because these contours are flat, especially when Flour = 400, we can infer robustness. Hence, when producing the cake mix industrially with the composition Flour 400g, Shortening 50g, and Eggpowder 100g, together with a recommendation on the box of using the temperature 200°C and time 40 min, sufficient robustness towards consumer misuse ought to be the result.
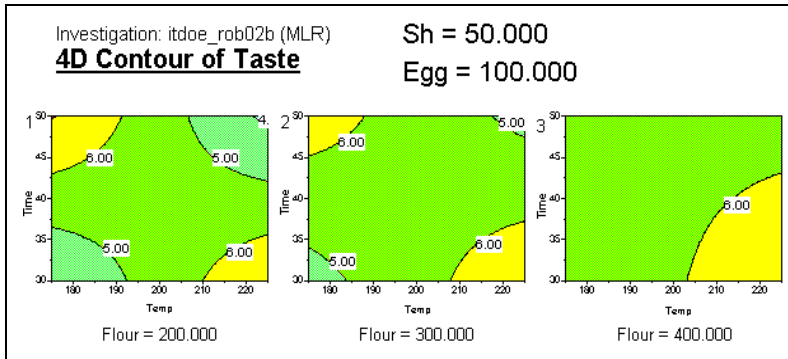
Figure 21.25: Response contour plots of Taste for varying levels of Time, Temperature and Flour, and with Shortening and Eggpowder fixed at 50 and 100g, respectively.

A second example – DrugD

We will now repeat the two analytical approaches to Taguchi planned experiments with an application taken from pharmaceutical industry, which we call DrugD. The background to the example is as follows. The manufacturer of a drug to be administered in tablet form wanted to test the robustness of the tablet for patients with varying conditions. Four factors were varied in an inner array setup to simulate a human stomach. These are seen in Figure 21.26 together with their settings.

Figure 21.26: (left) Factors varied in DrugD – classical analysis setup.
Figure 21.27: (right) Responses registered in DrugD – classical analysis setup.

Volume simulates the size of the stomach, Temperature whether the patient has fever or not, PropSpeed a calm or a stressed human being, and pH different levels of acidity in the gastrointestinal tract. The inner array used was a CCF design with 27 runs. This design was tested in 6 parallel baths, here labeled B1 – B6, and the measured response was the release after 1h of the active ingredient. In the written documentation the manufacturer declared

that after 1h the release should be between 20 and 40%. Hence, the experimental goal was to assess whether the variation in the release rates across the entire design was consistent with this claim.

In the classical analysis approach one would regard the qualitative Bath factor as the outer array, and use the average release and its standard deviation across the 6 baths as responses. These responses are summarized in Figure 21.27. Note the log-transform of the SD response. In the interaction analysis approach, one would combine the inner and outer array experiments in one single design, and fit a regression model with the qualitative factor also included. As seen in Figures 21.28 and 21.29 this means five factors and one response. The combined worksheet then has 27*6 =162 rows. In the following we will contrast the results obtained with the classical and the interaction analysis approaches.

Figure 21.28: *(left) Factors varied in DrugD – interaction analysis setup.*
Figure 21.29: *(right) Responses registered in DrugD – interaction analysis setup.*

DrugD – The classical analysis approach

The simultaneous analysis of the average release, denoted 1h, and its standard deviation, denoted SD1, using a full quadratic model, gave the modelling results rendered in Figure 21.30. After some model refinement, that is, exclusion of 6 non-significant cross- and square terms, we obtained two refined models with performance statistics as displayed in Figure 21.31.
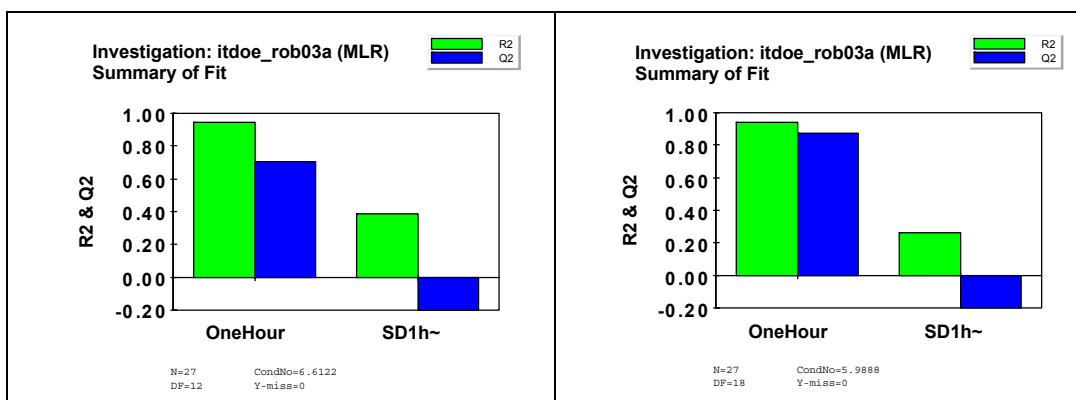
Figure 21.30: *(left) Summary of fit of initial models.*
Figure 21.31: *(right) Summary of fit of revised models.*

As seen from Figure 21.31, the model refinement resulted in an excellent model for 1h, but no model at all for SD1. This suggests that the variation in the four factors significantly affect the average release, but not the standard deviation. We can see from Figure 21.32 that

all factors but Volume influence the average release. In contrast, Figure 21.33 indicates that there are no significant effects with regard to the standard deviation. Hence, the latter response is robust to changes in the factors.
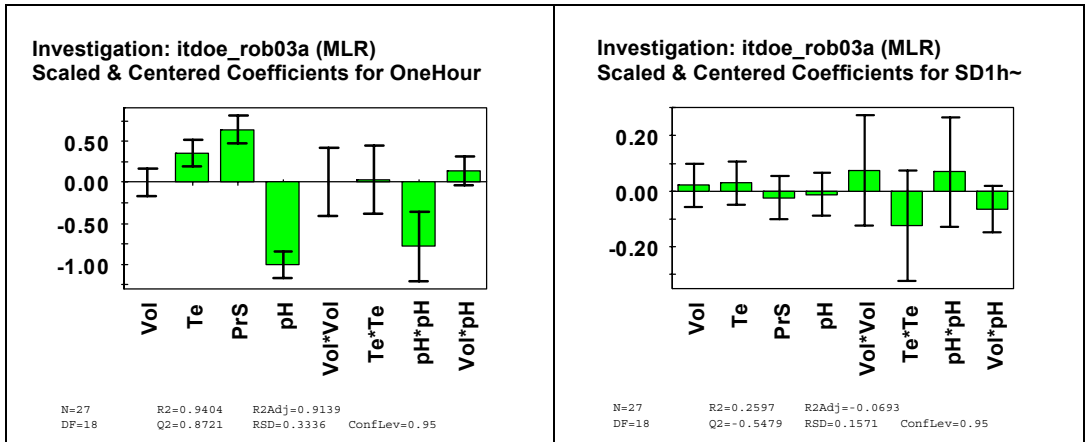
Figure 21.32: (left) Regression coefficients of revised model for 1h.
Figure 21.33: (right) Regression coefficients of revised model for SD1.

To better understand the features of the first response we created the response contour plot shown in Figure 21.34. From this figure it is easy to get the impression that the response 1h changes dramatically as a result of altered factor settings. However, this is not the case. We are fooled by the way this plot was constructed. A more appropriate plot is given in Figure 21.35. This is a response surface plot in which the z-axis, the release axis, has been re-expressed to go between 20 and 40%, which corresponds to the promised variation range. Now we can see the flatness of the response surface. Remarkably, the difference between the highest and the lowest measured values is as low as 4.1%. Hence, we can conclude that the average release response is robust, because it is inside the given specification.
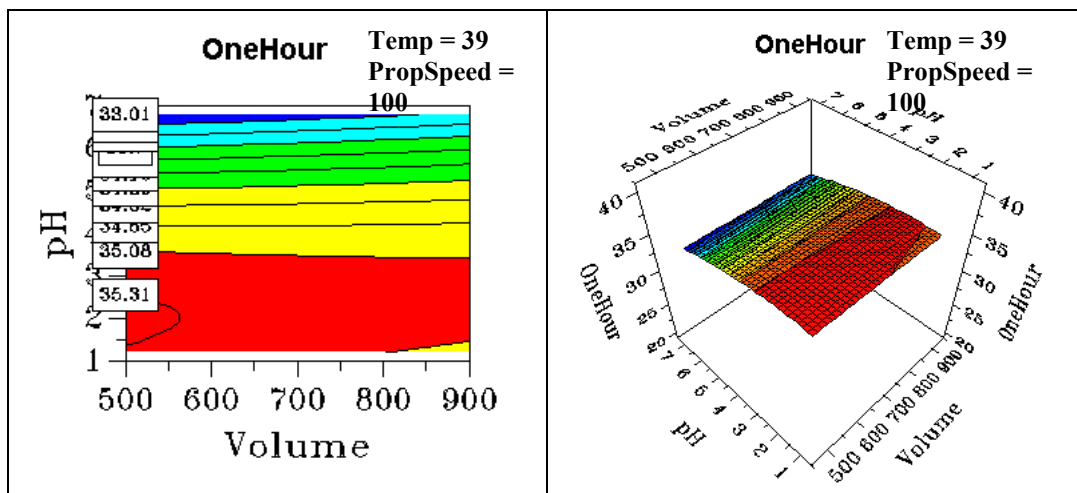
Figure 21.34: *(left) Response contour plot for 1h.*
Figure 21.35: *(right) Response surface plot for 1h.*

DrugD – The interaction analysis approach

For comparative purposes we now turn to the interaction analysis approach of the DrugD data. Figure 21.36 is a replicate plot over the entire set of 162 experiments, which shows the robustness of the tablet. In this case, we fitted a partially quadratic model, and, after some model refinement a model with $R^2 = 0.79$ and $Q^2 = 0.74$ was obtained. The summary of fit plot is provided in Figure 21.37.
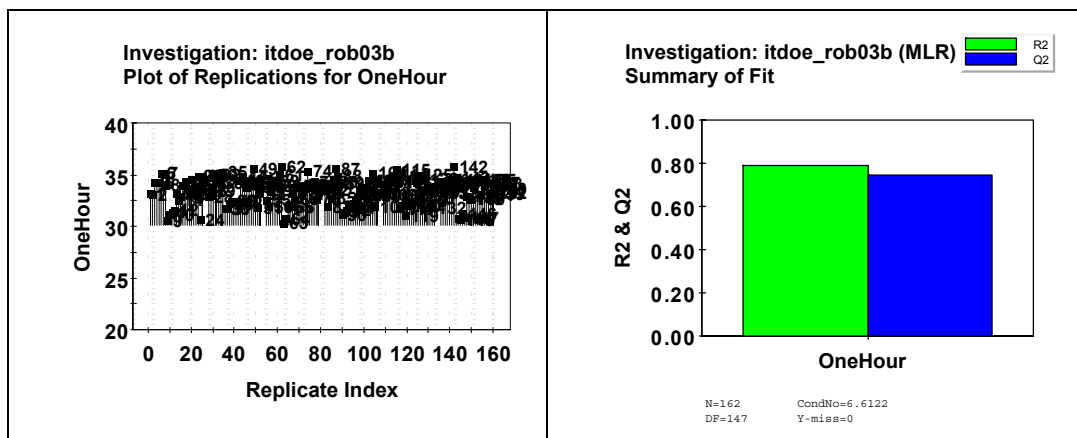
Figure 21.36: *(left) Replicate plot of 1h.*
Figure 21.37: *(right) Summary of fit of 1h.*

The N-plot of residuals in Figure 21.38 reveals no strange experiments, and the ANOVA table listed in Figure 21.39 also suggests that the model is adequate. The model itself is given in Figure 21.40. Because there were no strong model terms influencing the standard deviation in the foregoing analysis, we would not expect any significant interactions among the inner array factors and the Bath factor. This is indeed the case.
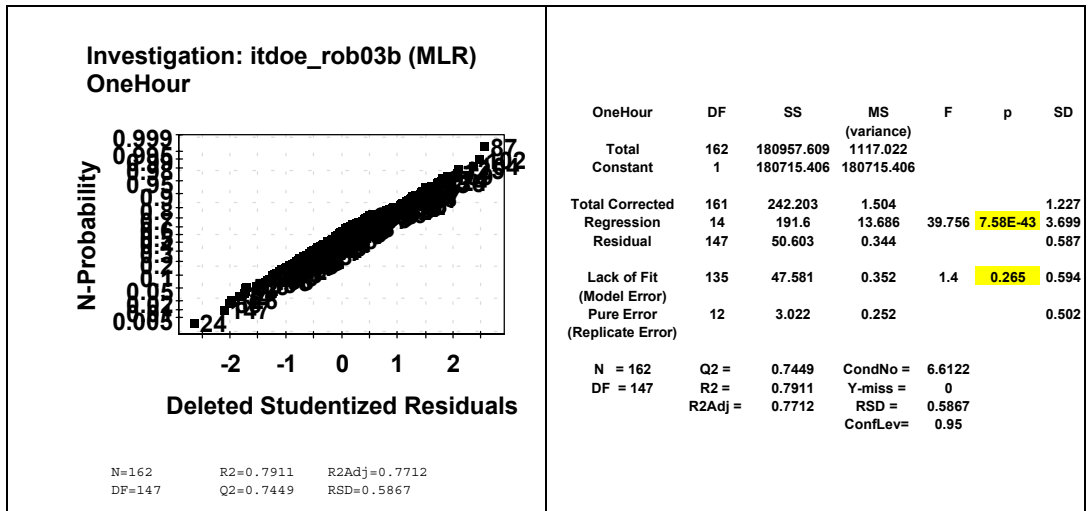
Figure 21.38: *(left) N-plot of residuals of 1h.*
Figure 21.39: *(right) ANOVA of 1h.*

Moreover, with the interaction analysis approach we are able to uncover one little piece of new information with regard to the 6 baths used. The regression coefficients suggest that with bath B2 the experimenters might obtain slightly higher numerical results than with the other five baths. Observe, though, that this effect of B2 is statistically weak and must not be over-interpreted. We show in Figure 21.41 an alternative version of Figure 21.40, in which the vertical scale is re-expressed to go between –5 and +5. Any model term of large magnitude in this plot would constitute a serious problem for achieving the experimental goal. This is not the case, and hence we understand that the effect of B2 is not large. However, the recognition of this small term is important for further fine-tuning of the experimental equipment.
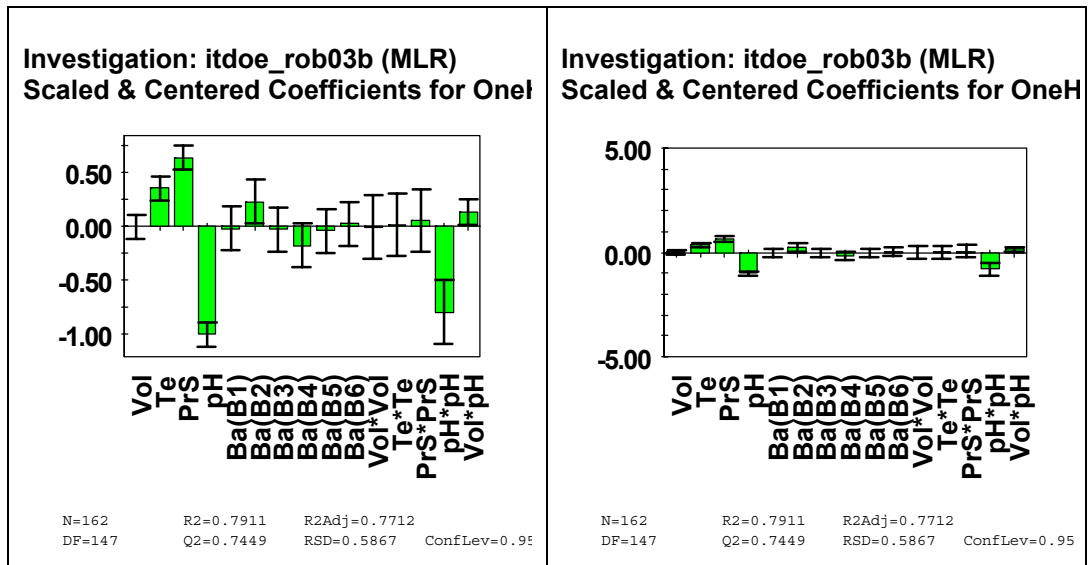
Figure 21.40: *(left) Regression coefficient plot of model for 1h.*
Figure 21.41: *(right) Same as Figure 21.40 but with a re-expressed y-axis.*

An additional element of robust design

Sometimes it is of interest to perform a robust design test where it is less relevant to differentiate between controllable and uncontrollable factors, and more appropriate to distinguish among factors which are *expensive* and *inexpensive* to vary. We will now do a thought experiment to illustrate this situation. Let us assume that we want to increase the durability and reliability of drills to be used for drilling stainless steel. Further, assume that our goal is to measure the lifetime of a drill under different conditions. This is a response which is cheap to measure and we wish to maximize it. In this case, we have two types of factors, those that are expensive to vary, and those that are inexpensive to change. Examples belonging to the costly category are factors related to the features of the drill, such as, diameter, length, and geometry. Examples of the cheaper type are factors related to the machine conditions, such as, cutting speed, feed rate, cooling, and so on.
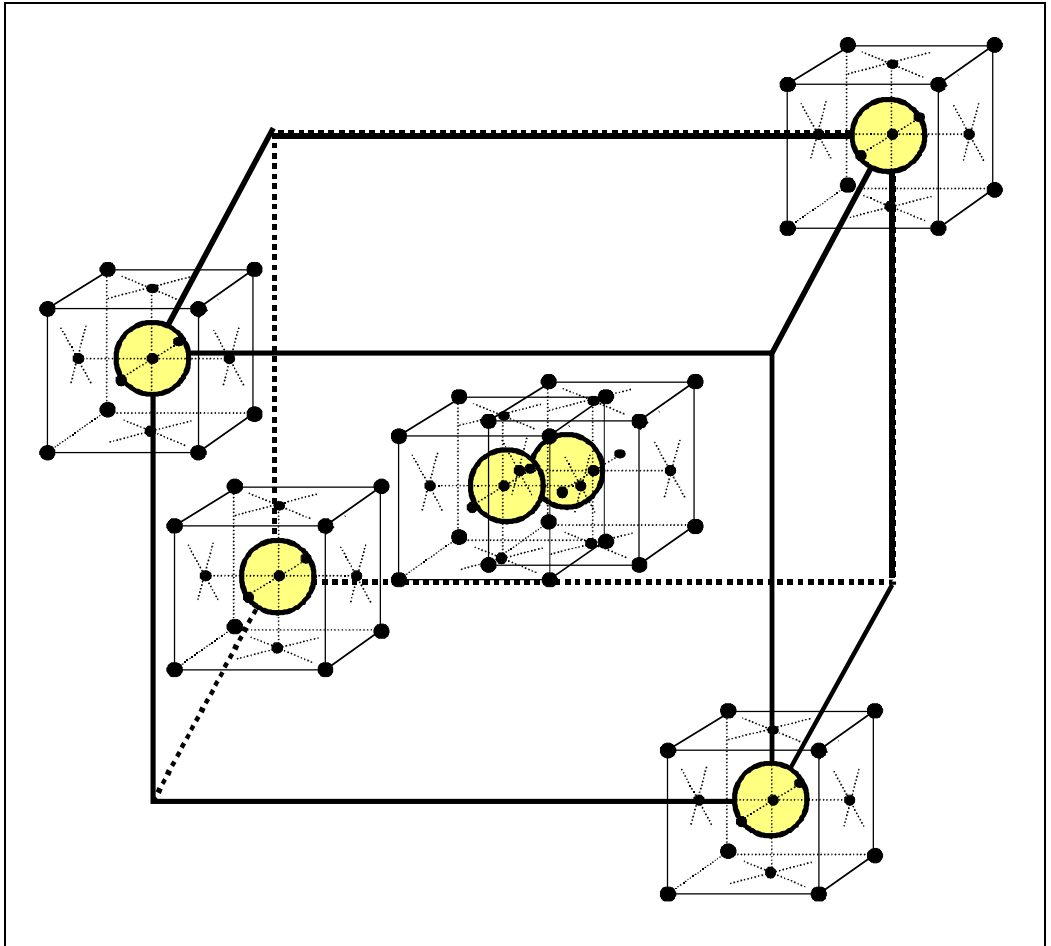
Figure 21.42: *A combined experimental protocol for the testing of expensive and inexpensive factors in drilling.*

In this kind of application it is relevant to try to minimize the number of alterations made with the most expensive factors, without sacrificing the quality of the work. Thus, to adequately address this problem we propose the following: A combined experimental design of the type shown Figure 21.42 is created, where the inner array is a medium resolution fractional factorial design laid out in the expensive factors, and the outer array is a composite design defined for the inexpensive factors. For instance, if we have three expensive and three inexpensive factors, we could lay out a combined design in 102 experiments, or 17 experiments per drill. This would be one way of assessing the robustness of a drill towards different customer practices, without draining the available test resources.

Quiz

Please answer the following questions:

What is a design factor in the Taguchi approach? A noise factor?

How are design and noise factors varied in a Taguchi design?

How does the classical analysis approach work?

What are its advantages and disadvantages?

Why is it usually necessary to log-transform a standard deviation response?

How does the interaction analysis approach work?

What are its advantages and disadvantages?

Which procedure may be used for robust design of expensive and inexpensive factors?

Summary

In this chapter, we have described the Taguchi approach to robust design. Briefly, in this approach a distinction is made between design factors and noise factors. The former are easy-to-control and correspond to the factors that we normally vary in a statistical experimental design. The latter are hard-to-control and correspond to factors that in principle should have no effect on the responses. Taguchi proposed that design factors and noise factors be co-varied together by means of the inner and outer array system. Experimental data generated with the inner and outer array planning can be analyzed in two ways. In the classical analysis approach, the average response and the standard deviation are used as response variables. One is then interested in finding out which design factors mainly affect the signal, the average, and which affect the noise, the standard deviation. Note that with this approach no effect estimates are obtainable for the noise factors. In the alternative approach, the interaction analysis approach, a combined design is created, which includes all experiments of the inner and outer arrays. This makes it possible to fit a regression model in which model terms of the noise factors are included. It also enables the detection of single deviating experiments. It is important to realize that what in the classical analysis approach were design factor effects on the standard deviation response, now correspond to noise-design factor interactions. Recall that we encountered a strong three-factor interaction in the CakeMix study.

22 Models of variability: Distributions (Level 3)

Objective

The objective of this chapter is to provide an introduction to some basic univariate statistical concepts, which are useful for the evaluation DOE data. We will first focus on the concept of variability, which was partly introduced in Chapter 2. In order to deal with variability, we will need models of variability. Such models are called distributions, and the first part of this chapter is devoted to the discussion of some common distributions. The distributions that will be treated are the normal distribution, the t-distribution, the log-normal distribution and the F-distribution. These distributions give rise to two diagnostic tools which are useful in DOE, the confidence interval and the F-test. Here, we will demonstrate how to compute confidence intervals for regression coefficients.

Models of variability - Distributions

Suppose we have measured the yield of a product ten times, and that these measurements were registered under identical experimental conditions. Then we might end up with a situation resembling the data displayed in Figure 22.1. Apparently, these data vary, despite the fact that they were obtained using identical conditions. The reason for this variation is that every measurement and every experiment is influenced by noise. This happens in the laboratory, in the pilot-plant, and in full-scale production.

Ten measurements of yield, under identical conditions

- Data vary,
- but vary in a *limited interval,*
- and with a central tendency

average:
$$\overline{x} = \sum (x_i) / N = 94.4$$

standard deviation:
$$S = \sqrt{\sum (x_i - \overline{x})^2 / (N-1)} = 1.56$$
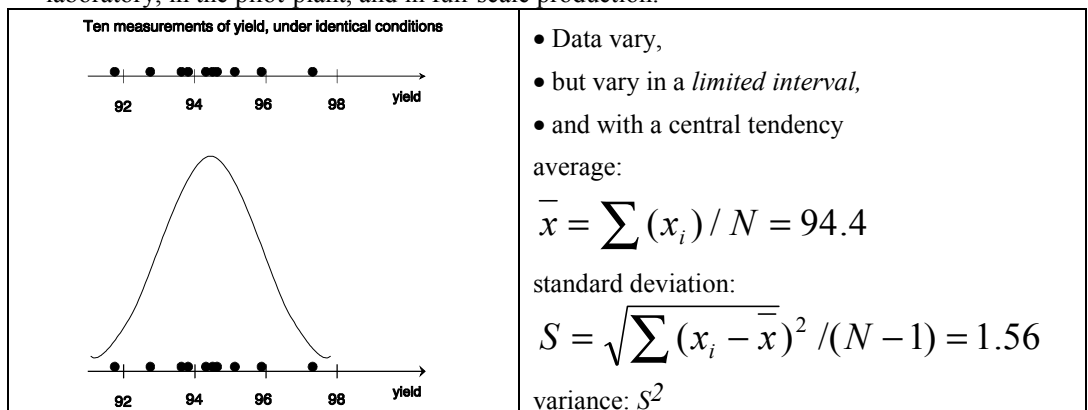
variance: S^2

Figure 22.1: (left) Data vary even when acquired under identical conditions.
Figure 22.2: (right) The average and the standard deviation of the data displayed in Figure 22.1.

In order to draw correct conclusions, an experimenter must be able to determine the size of the experimental variation. To accomplish this, however, the experimenter needs tools to handle variability. Such tools, or models, of variability exist and they are called *distributions*. Figure 22.1 illustrates what a distribution might look like in the context of the 10 replicated experiments. We can see that the data vary in a limited interval and with a tendency to group around a central value. The *size* of this interval, usually measured with the *standard deviation*, and the *location* of this central value, usually estimated with the *average*, may be used to characterize the properties of the occurring variability. Figure 22.2 lists the standard deviation and the average value of the 10 experiments.

The normal distribution

The most commonly used theoretical distribution is the Normal or Gaussian distribution, which corresponds to a bell-shaped probability curve. A schematic example for a variable x is provided in Figure 22.3. In this probability graph, the y-axis gives the probability count and the x-axis the measured values of the studied variable. For convenience, this variable may be re-expressed to become a *standardized normal deviate* z, which is a variable standardized such that the mean is 0 and the standard deviation 1. Figure 22.4 shows how this standardization is conducted.

$$z = (x - \mu) / \sigma;$$
- μ is population mean
- σ is population standard deviation
- z has mean = 0 and SD = 1
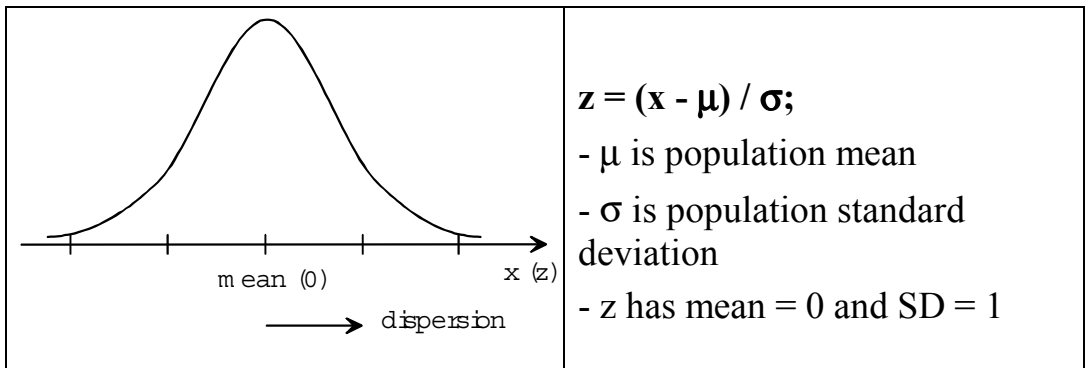
Figure 22.3: *(left) Normal probability curve of a variable x or a standardized normal variate z.*
Figure 22.4: *(right) How to standardize a variable to get average 0 and standard deviation 1.*

One interesting property of the normal distribution is that the area under the curve corresponds to a probability level, and that the area under the whole curve is 1 or 100% probability. This property means that we may now conduct an assessment of the probability that |z| will not exceed a given reference value z_0. Such probability levels are tabulated in Figure 22.5.

Probability, P, that |z| will not exceed a given value, z₀

z_0	P
0.5	0.383
1	0.683
1.5	0.866
2	0.955
2.5	0.988
3	0.9973
4	0.9999

Figure 22.5: Probability, P, that |z| will not exceed a given value, z₀

It is seen that a normally distributed variable will seldom give a value of |z| greater than 3. An alternative and perhaps easier way to understand this phenomenon consists of the following: Consider the mean value of 0. From this mean and out to $-1\,\sigma$ and $+1\,\sigma$ an area corresponding to 68.3% of the total area under the curve is covered. Similarly, in the range mean $\pm\,2\,\sigma$'s the covered area is 95.5% of the total area, and when using the range mean $\pm\,3$ σ's 99.7% of the total area is covered. Hence, when working with the normal distribution, almost all data are found in the interval mean $\pm\,3\,\sigma$.

In summary, the normal distribution makes it possible for the experimentalist to quantify variability. We will now describe how such information may be used for statistical inference.

The t-distribution

A common statistical operation is to calculate the *precision* of some estimated statistic, for instance, a mean value. Usually, such a precision argument is known as a *confidence interval*. One term contributing to such a precision estimate is the standard deviation of the data series distribution. If the form of this distribution is well-established, and if its true mean, μ, and true standard deviation, σ, are both known, it is possible to make straightforward probability statements about the mean value of a number of observations. In most practical circumstances, however, the true standard deviation σ is not known, and only an estimate, s, of σ, based on a limited number of degrees of freedom, is available.

Whenever the standard deviation s is estimated from the data, a distribution known as the *t-distribution* has to be employed. A simplified illustration of the t-distribution is given in Figure 22.6. As seen, the t-distribution is more peaked than the normal distribution and has heavier tails. These special features of the t-distribution arise because the estimated standard deviation, s, is itself subject to some uncertainty. Observe that with sufficiently many degrees of freedom, say, more than 30, the uncertainty in s is comparatively small, and the t-distribution is practically identical to the normal distribution. We will now examine how the t-distribution may be utilized to accomplish a precision estimate of a mean value.
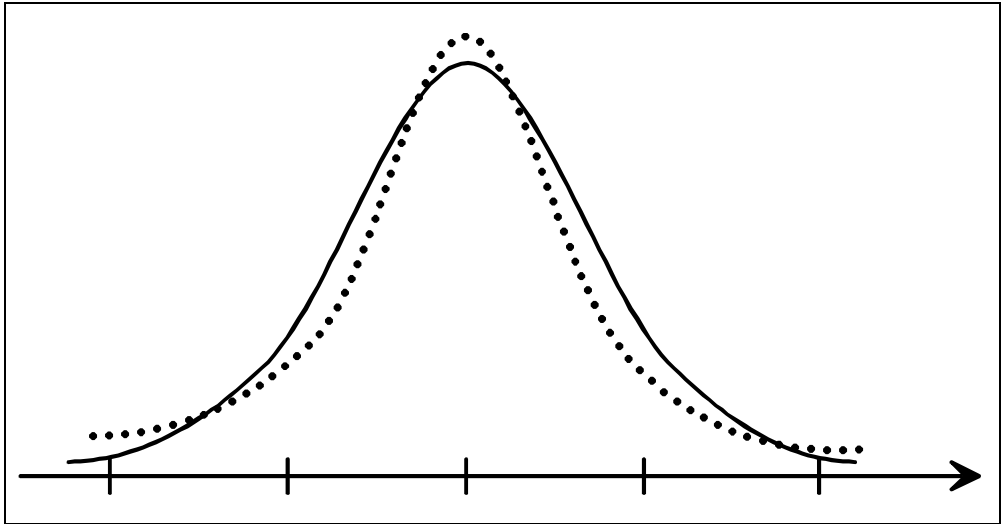
Figure 22.6: A comparison of the normal distribution and the t-distribution.

Confidence intervals

When estimating a mean value we would like to know the significance of this parameter, that is, we would like to know not only the estimated value of the statistic, but also how precise it is. In other words, we want to be able to state some reference limits within which it may reasonably be declared that the true value of the statistic lies. Such statements may assert that the true value is unlikely to exceed some upper limit, or that it is unlikely to be less than some lower limit, or that it is unlikely to lie outside a pair of limits. Such a pair of limits is often known as *confidence limits*, or a *confidence interval*, and is just as important as the estimated statistic itself. The degree of confidence used is usually set at 95%, but higher or lower levels may be chosen by the user.

We will now illustrate how to compute a confidence interval of a mean value. Consider the series of ten replicated experiments. The mean value of this series is 94.4 and the standard deviation 1.56. In order to compute the confidence interval around this mean, one proceeds as depicted in Figure 22.7. The first step is to calculate the standard error, SE, which is the standard deviation of the mean. Hence, SE is expressed in the original measurement unit and constitutes a dispersion estimate around the mean. Secondly, in order to convert the standard error into a confidence expression, we need a reference t-value acquired from an appropriate t-distribution. The numerical value of t reflects the size of the data set, that is, the available degrees of freedom, and the chosen level of confidence. By inspecting a table of t-values, we can see that the relevant t-value for nine degrees of freedom is 2.262 if a 95% confidence interval is sought. Observe that the t-distribution is listed as a single-sided argument, whereas the confidence interval is a double-sided argument, and hence a t-value of confidence level 0.025 is used. Finally, in the third step, the SE is multiplied by t and attached to the estimated mean value. Figure 22.8 shows how the confidence interval of the mean value may be interpreted graphically. The numerical range from 93.3 to 95.5 corresponds to the interval within which we can be 95% confident that the true mean will be found.

SE (standard error = SD of the mean) = $1.56/\sqrt{10}$ = 0.493

t (9, 0.025) = 2.262; from a t-table with 9 (10-1) degrees of freedom

The 95% confidence interval for our average is: 94.4 $\pm$ (2.262* 0.493) = 94.4 $\pm$ 1.1
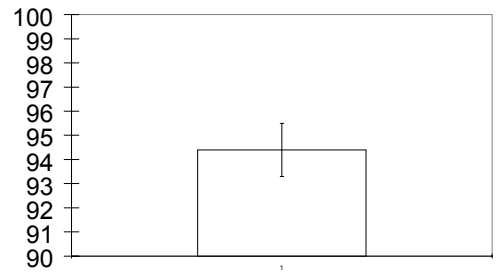
Figure 22.7: *(left) How to compute a confidence interval of a mean value.*
Figure 22.8: *(right) How to display a confidence interval of a mean value.*

Confidence intervals of regression coefficients

The use of a mean value and its confidence interval closely resembles how the statistical significance of regression coefficients is assessed. We recall from Chapter 2, that a regression coefficient is estimated in terms of averaging several differences in response values between high and low levels of the varied factors. Hence, regression coefficients may be interpreted as mean values, and their precision understood through the calculation of 95% confidence intervals. As an example, consider the regression coefficients of the CakeMix example plotted in Figure 22.9.
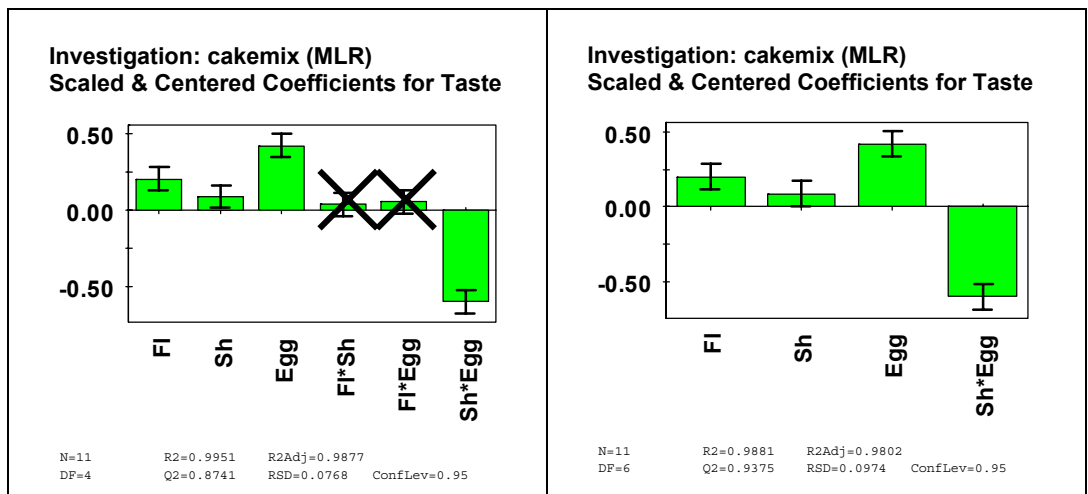
Figure 22.9. *(left) Confidence intervals of regression coefficients – CakeMix example.*
Figure 22.10: *(right) Confidence intervals and regression coefficients after model refinement.*

The appearance of the confidence intervals reveal that the value of zero is plausible for two of the interaction terms. Because of this, we can conclude that these terms are not statistically significant, and accordingly they can be removed from the model. The result after the removal of these terms and refitting of the model is displayed in Figure 22.10. It is interesting to note that Q^2 has increased from 0.87 to 0.94, which vindicates the model pruning.

The example outlined indicates that the confidence interval of a regression coefficient is an informative tool for obtaining better models. The computation of such confidence intervals is slightly more complex than for an ordinary mean value. Figure 22.11 shows how this is done for well-designed data, and the essence of the given expression is explained in the statistical appendix. At this stage, it suffices to say that three terms are involved in the computation of confidence intervals, and that the sharpest precisions are obtained with (i) a good design with low condition number, (ii) a good model with low residual standard deviation, and (iii) sufficiently many degrees of freedom.

$$\pm\sqrt{(X'X)^{-1}}*RSD*t(\alpha/2, DF_{resid})$$

Figure 22.11: *How to compute confidence intervals of regression coefficients.*

Good model is obtained with:

- a good design of low condition number
- a good model with low RSD
- sufficiently many degrees of freedom

The log-normal distribution

We will now consider another distribution which is commonly encountered in DOE, the log-normal distribution. Sometimes it will be found that the distribution of a variable departs from normality, and it would be desirable if, by a simple transformation of the variable, an approximately normal distribution could be obtained. The principles for this are illustrated in Figures 22.12 and 22.13.
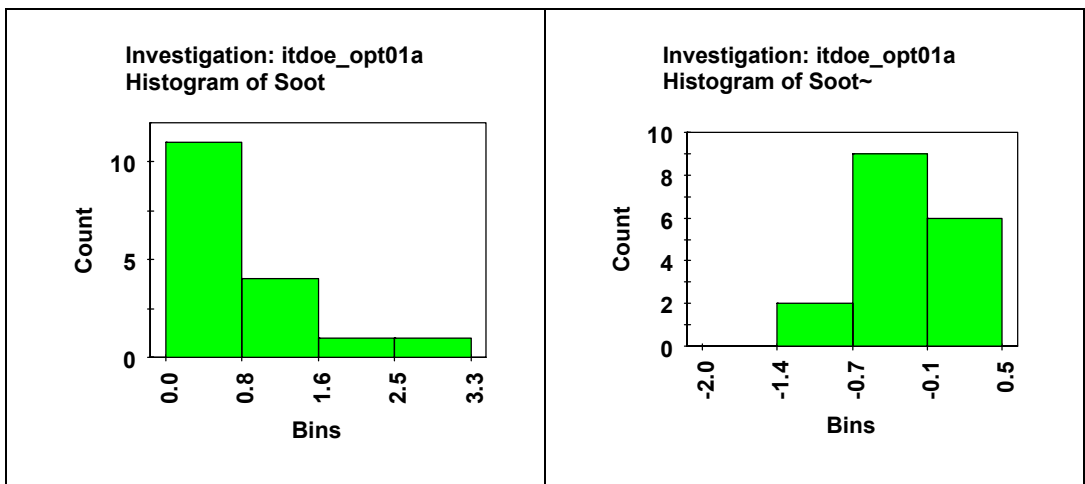
Figure 22.12: *(left) Histogram of response Soot of truck engine data.*
Figure 22.13: *(right) Histogram of log-transformed response.*

Figure 22.12 shows the histogram of the Soot response of the truck engine application. The appearance of this histogram suggests that Soot adheres to the log-normal distribution. This distribution is characterized by a heavy tail to the right, that is, a few large numerical values and a majority of low measured values. By applying a logarithmic transformation to the raw data of Soot, a distribution which is much closer to normality is obtained (Figure 22.13). This transformed variable is better suited for regression analysis, and will make the modelling easier. Other examples of transformations, which are often used for transferring a skew distribution into an approximate normal distribution, are the square root, the fourth root, and the reciprocal of a variable.

The F-distribution

The last distribution which will be considered here is the *F-distribution*. In contrast to the t-distribution, which is used in conjunction with statistical testing of, for instance, mean values, the F-distribution is used for comparing variances with each other. It is possible to test whether two estimated variances differ significantly. This testing is carried out by forming the *ratio* between two compared variances. The larger variance is used in the numerator and the smaller variance in the denominator. When the underlying measured values are normally distributed, a variance ratio formed in this manner is distributed according to a distribution known as the F-distribution. Further, this ratio only depends on the degrees of freedom associated with the two variance estimates. In DOE, the F-test is ubiquitous in analysis of variance, ANOVA, which is explained in Chapter 23.

Quiz

Please answer the following questions:

What is a distribution?

What is the normal distribution?

When is the t-distribution appropriate?

What is the essence of a confidence interval?

How can a confidence interval be used for model refinement?

What is the log-normal distribution?

Which transformations can be used to obtain approximately normally distributed data?

Which distribution is used for comparing variances?

Summary

In this chapter, we have considered variability and models for variability. Such models are called distributions, and they enable uncertainty to be handled in a rational manner. Initially, the most commonly used distribution, the normal distribution, was scrutinized. Often, however, in DOE, the number of runs in an experimental plan is rather small, rendering statistical reasoning based on the normal distribution inappropriate. As a consequence, much use is made of the t-distribution, for example in the estimation of confidence intervals of regression coefficients. Another common distribution is the log-normal distribution. Response variables whose numerical values extend over more than one order of magnitude often display a log-normal distribution. This skewness can be altered by using an

appropriate transformation, like the logarithmic transformation. Other common transformations are the square root, the fourth root, and the reciprocal of the variable. Finally, we discussed the F-distribution, which is used to make assessments between two compared variances. Variance testing is done in analysis of variance, ANOVA.

23 Analysis of variance, ANOVA (Level 3)

Objective

Analysis of variance, ANOVA, is used as a basis for regression model evaluation. This technique was briefly introduced in Chapter 11. It is the objective of the current chapter to provide more details of ANOVA and its use in connection with regression analysis. Since much of the statistical testing in ANOVA is based on the F-test, we will also explain this test.

Introduction to ANOVA

Multiple linear regression, MLR, is based on finding the regression model which minimizes the residual sum of squares of the response variable. We have seen in the earlier part of the course, that with designed data it is possible to add terms to or remove terms from an MLR model. The question which arises then is *how can we be sure that any revised model is better than the model originally postulated*? Fortunately, we can use analysis of variance, ANOVA, R^2/Q^2 and residual plots to shed some light on this question. ANOVA makes it possible to formally evaluate the performance of alternative models. We will now study how this is carried out.

Consider the CakeMix application, and recall that the performed experiments support the interaction model $y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_{12}x_1x_2 + \beta_{13}x_1x_3 + \beta_{23}x_2x_3 + \varepsilon$. We found in Chapter 10 that when applying MLR to the CakeMix data, the model shown in Figure 23.1 was the result. However, we also found that two model terms could be removed from the model and the model thus refined. The revised model is shown in Figure 23.2.

Investigation: cakemix (MLR)
Scaled & Centered Coefficients for Taste

0.50

0.00

-0.50

Fl Sh Egg Fl*Sh Fl*Egg Sh*Egg

N=11   R2=0.9951   R2Adj=0.9877
DF=4   Q2=0.8741   RSD=0.0768   ConfLev=0.95

Investigation: cakemix (MLR)
Scaled & Centered Coefficients for Taste

0.50

0.00

-0.50

Fl Sh Egg Sh*Egg

N=11   R2=0.9881   R2Adj=0.9802
DF=6   Q2=0.9375   RSD=0.0974   ConfLev=0.95
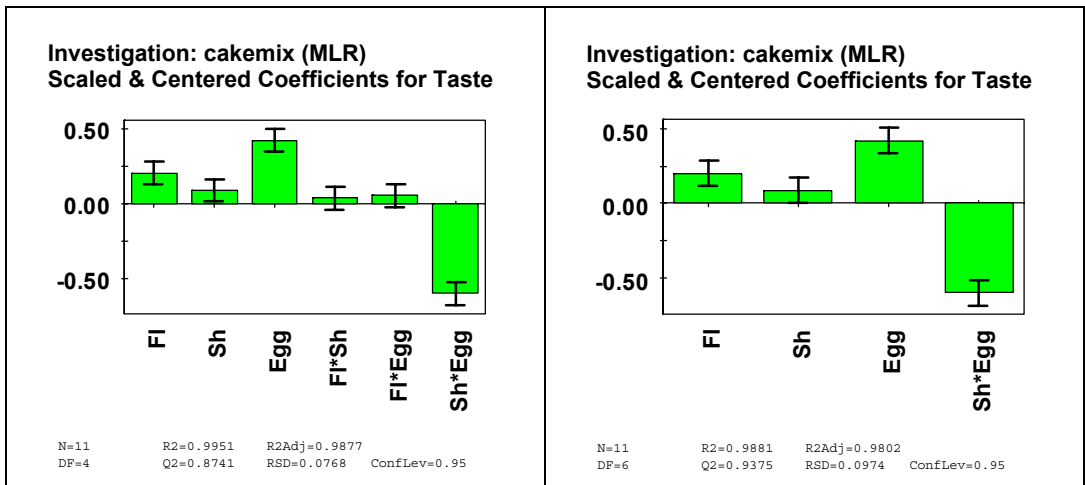
Figure 23.1: (left) Original model of Taste.
Figure 23.2: (right) Revised model of Taste.

Now, it is of interest to compare these models and this may be done with ANOVA. ANOVA is based on partitioning the total variation of a selected response into one part due to the regression model and another part due to the residuals. At times, when replicated experiments are available, ANOVA also decomposes the residual variation into one part related to the model error and another part linked to the replicate error. Subsequently, the numerical sizes of these variance estimates are formally compared by means of F-tests. How to interpret such tables is explained below.

Taste	DF	SS	MS (variance)	F	p	SD
Total	11	247.205	22.473			
Constant	1	242.426	242.426			
Total Corrected	10	4.778	0.478			0.691
Regression	6	4.755	0.792	134.469	1.45E-04	0.89
Residual	4	0.024	0.006			0.077
Lack of Fit (Model Error)	2	0.016	0.008	2.244	0.308	0.09
Pure Error (Replicate Error)	2	0.007	3.63E-03			0.06

N = 11   Q2 = 0.8741   CondNo = 1.1726
DF = 4   R2 = 0.9951   Y-miss = 0
           R2Adj = 0.9877 RSD = 0.0768

Taste	DF	SS	MS (variance)	F	p	SD
Total	11	247.205	22.473			
Constant	1	242.426	242.426			
Total Corrected	10	4.778	0.478			0.691
Regression	4	4.721	1.18	124.525	6.68E-06	1.086
Residual	6	0.057	0.009			0.097
Lack of Fit (Model Error)	4	0.05	0.012	3.413	0.239	0.111
Pure Error (Replicate Error)	2	0.007	3.63E-03			0.06

N = 11   Q2 = 0.9375   CondNo = 1.1726
DF = 6   R2 = 0.9881   Y-miss = 0
           R2Adj = 0.9802 RSD = 0.0974
                           ConfLev= 0.95

Figure 23.3: (left) ANOVA of original model for Taste.
Figure 23.4: (right) ANOVA of refined model for Taste.

ANOVA – Regression model significance test

In ANOVA, a common term is *sum of squares*, SS. Sums of squares are useful to quantify variability, and can be decomposed into smaller constituents with ANOVA. In ANOVA, the first decomposition is $SS_{total\ corrected} = SS_{regression} + SS_{residual}$. The first term, $SS_{total\ corrected}$, is the total variation in the response, corrected for the average. With least squares analysis we want to create a mathematical model, which can describe as much as possible of this total

variation. The amount of variation that we can model, "describe", is given by the second term, $SS_{regression}$. Consequently, the amount of variation that we can *not* model is given by the third term, $SS_{residual}$.

Usually, a model is good when the modellable variation, $SS_{regression}$, is high and the unmodellable variation, $SS_{residual}$, is low. It is possible to formulate a formal test to check this. In doing so, one first converts these two *variation* estimates into their *mean squares (variance)* counterparts, that is, $MS_{regression}$ and $MS_{residual}$. This is accomplished by dividing the SS estimates with the corresponding degrees of freedom, DF. Subsequently, the sizes of these two variances are formally compared by an F-test. This is accomplished by forming the ratio $MS_{regression}/MS_{residual}$ and then retrieving the probability, p, that these two variances originate from the same distribution. It is common practice to set $p = 0.05$ as the critical limit. Figures 23.3 and 23.4 show that the relevant p-value for the original CakeMix model is 1.45E-04 and for the revised model 6.68E-06. These values are well below 0.05, indicating that each model is good. However, the p-value is lower for the revised model and this suggests that the model refinement was appropriate. In conclusion, the variance explained by either version of the model is significantly larger than the unexplained variance.

ANOVA – Lack of fit test

Furthermore, in ANOVA, a second decomposition of sums of squares may be made. This decomposition is done according to $SS_{residual} = SS_{model\ error} + SS_{replicate\ error}$. Thus, the unmodellable variation, $SS_{residual}$, has two components, one arising from the fact that the model is imperfect, the model error, and one arising from the fact that there is always variation when doing replicated experiments, the replicate error. In the computation of the replicate error one considers the replicated experiments and their deviation around the local replicate mean value. Once the replicate error has been determined it is subtracted from the residual sum of squares, and the remainder then corresponds to the model error.

In the ideal case, the model error and the replicate error are small and of similar size. Whether this is the case may be formally tested with an F-test. In this test, the sizes of the two variances $MS_{model\ error}$ and $MS_{replicate\ error}$ are compared by forming their ratio and F-testing this ratio. We can infer from the ANOVA table in Figure 23.3 that the model error of the original model is of the same magnitude as the replicate error, because the p-value 0.308 is *larger* than the critical reference value of 0.05. Hence, our model has small model error and good fitting power, that is, shows no *lack of fit*. As seen Figure 23.4, the p-value for the alternative model is slightly lower, 0.239, indicating that the original model is preferable to the revised model. However, the important fact is that both models pass the lack of fit test, and the fact that the p-value is slightly lower for the revised model is of only marginal relevance.

Summary of ANOVA

We have now seen that with ANOVA pairs of variances are formally compared by conducting F-tests. The first F-test, comparing modellable and unmodellable variance, is satisfied when p is < 0.05. The second F-test, comparing model and replicate errors with each other, is satisfied when p is > 0.05. The second F-test is called the *lack of fit* test and is an important diagnostic test, but it cannot be carried out if replicates are lacking.

In addition, the ANOVA table contains other statistical measures, which are useful in model evaluation. One parameter is the *explained variation*, R^2, which is calculated as $1 - SS_{residual}/SS_{total\ corrected}$, and lies between 0 and 1. This parameter is the classical quantity used for model evaluation. Unfortunately, R^2 is very sensitive to the degrees of freedom, and by including more terms in the model it is possible to fool R^2 and make it arbitrarily close to 1. Recognizing this deficiency of R^2, one may choose to work with the *explained variance*, R^2_{adj}, which is a goodness of fit measure adjusted for degrees of freedom. The adjusted R^2 is computed as $1 - MS_{residual}/MS_{total\ corrected}$, and this parameter is always lower than R^2.

A third parameter often employed in this context is the *predicted variation*, Q^2, which is computed in the same way as R^2, the only difference being that $SS_{predictive\ residual}$ is used in place of $SS_{residual}$. Q^2 estimates the predictive power of a model, and is therefore of primary interest in regression modelling. We can see in Figures 23.3 and 23.4, that R^2 and R^2_{adj} have decreased a little as a result of the model revision. Normally, R^2 decreases when less useful model terms are removed, whereas R^2_{adj} should essentially remain unchanged. Also, if irrelevant model terms are omitted from the model, it is expected that Q^2 should increase. We can see that Q^2 has increased from 0.87 to 0.94, that is, we have obtained a verification that the model revision was appropriate.

The F-test

We will end this chapter by considering the F-test. The F-test compares the ratio of two variances and returns the probability that these originate from the same distribution, that is, the probability that these are not significantly different. As an example, consider the upper F-test of the original CakeMix model (Figure 23.3). The larger variance, $MS_{regression}$, is 0.792, and the smaller variance, $MS_{residual}$, is 0.006, and they have 6 and 4 degrees of freedom, respectively. The F-value is found by forming the ratio of these two variances, which is 134.469. Figure 23.5 shows an F-table pertaining to $p = 0.05$, and it shows that with 6 and 4 degrees of freedom the critical F-value is 6.2. Because the obtained F-value of 134.469 is larger than the critical F-value, the conclusion is that the two variances are unequal and are not drawn from the same distribution, that is, they are significantly different.

	1	2	3	4	5	6	7	8	9	10	11	12
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.0	243.9
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4
3	10.1	9.6	9.3	9.1	9.0	8.9	8.9	8.8	8.8	8.8	8.8	8.7
4	7.7	6.9	6.6	6.4	6.3	6.2	6.1	6.0	6.0	6.0	5.9	5.9
5	6.6	5.8	5.4	5.2	5.1	5.0	4.9	4.8	4.8	4.7	4.7	4.7
6	6.0	5.1	4.8	4.5	4.4	4.3	4.2	4.1	4.1	4.1	4.0	4.0
7	5.6	4.7	4.3	4.1	4.0	3.9	3.8	3.7	3.7	3.6	3.6	3.6
8	5.3	4.5	4.1	3.8	3.7	3.6	3.5	3.4	3.4	3.3	3.3	3.3
9	5.1	4.3	3.9	3.6	3.5	3.4	3.3	3.2	3.2	3.1	3.1	3.1
10	5.0	4.1	3.7	3.5	3.3	3.2	3.1	3.1	3.0	3.0	2.9	2.9
11	4.8	4.0	3.6	3.4	3.2	3.1	3.0	2.9	2.9	2.9	2.8	2.8
12	4.7	3.9	3.5	3.3	3.1	3.0	2.9	2.8	2.8	2.8	2.7	2.7

Figure 23.5: *An excerpt of the $p = 0.05$ F-table used for assessing the significance of the CakeMix regression model.*

The p-value given in the ANOVA table (Figure 23.3) of 1.45E-04 states the real probability at which a critical reference value would be equal to the tested ratio. This is shown by the excerpt of the F-table corresponding to $p = 1.45E-04$ in Figure 23.6.

	3	4	5	6	7
3	514.1	499.2	489.9	483.4	478.7
4	149.9	142.5	137.8	**134.5**	132.0
5	74.0	69.2	66.1	63.9	62.4
6	47.1	43.4	41.1	39.5	38.3
7	34.5	31.5	29.5	28.2	27.2

Figure 23.6: *Excerpt of the p = 1.45E-04 F-table.*

Quiz

Please answer the following questions:

What is the basic intention of ANOVA?

Which two tests are carried out in ANOVA?

What is an F-test?

What is R^2? R^2_{adj}? Q^2?

Summary

Analysis of variance, ANOVA, is an important diagnostic tool in regression analysis. ANOVA partitions the total variation of a response variable into one component due to the regression model and another component due to the residuals. Furthermore, when replicated experiments are available, ANOVA also decomposes the residual variation into one part related to the model error and another part linked to the replicate error. Then, the numerical sizes of these variance estimates are formally compared by means of F-tests. The first F-test, comparing modellable and unmodellable variances, is satisfied when p is smaller than 0.05. The second F-test, comparing model and replicate errors, is satisfied when p exceeds 0.05. The second F-test is called the *lack of fit* test and is an important diagnostic test, but it cannot be carried out if replicates are lacking. In this context, we also examined the properties of the F-test. Moreover, ANOVA gives rise to the goodness of fit parameters R^2 and R^2_{adj}, which are called *explained variation* and *explained variance*, respectively. The former parameter is sensitive to degrees of freedom, whereas for the latter this sensitivity has been reduced, thereby creating a more useful parameter. In the regression analysis it is also possible to compute the goodness of prediction, Q^2, which is the most realistic parameter of the three. Observe that Q^2 is not directly given by ANOVA, but is derivable through similar mathematical operations.

24 PLS (Level 3)

Objective

In this chapter, the objective is to describe PLS. PLS is an alternative to MLR, which can be used for the evaluation of complex experimental design data, or even non-designed data. PLS is an acronym of partial least squares projections to latent structures, and is a multivariate regression method. We will begin this chapter with a polymer example containing 14 responses and show how PLS may be applied to this data set. This example will help us to position PLS appropriately in the DOE framework, and highlight its essential modelling features. We will then try to introduce PLS from a geometrical perspective, but also provide relevant equations. A statistical account of PLS is given in the statistical appendix.

When to use PLS

When several responses have been measured, it is useful to fit a model simultaneously representing the variation of all responses to the variation of the factors. This is possible with PLS, because PLS deals with many responses by taking their covariances into account. MLR is not as efficient in this kind of situation, because separate regression models are fitted for each response.

Another situation in which PLS is appealing is when the experimental design is distorted. PLS handles distorted designs more reliably than MLR, since MLR in principle assumes perfect orthogonality. A design distortion may, for instance, occur because one corner of the design was inaccessible experimentally, or because some critical experiment failed. Usually, a distorted design has a high condition number, that is, it has lost some of its sphericity. As a rough rule of thumb, as soon as the condition number exceeds 10, it is recommended to use PLS. MLR should not be used with such high condition numbers, because the interpretability of the regression coefficients breaks down. Some regression coefficients become larger than expected, some smaller, and some may even have the wrong sign.

A third argument in favor of PLS is when there are missing data in the response matrix. MLR cannot handle missing data efficiently, and therefore each experiment for which data are missing, must be omitted from the analysis. PLS can handle missing data, as long as they are missing in a random fashion. In summary, PLS is a pertinent choice, if (i) there are several correlated responses in the data set, (ii) the experimental design has a high condition number, or (iii) there are small amounts of missing data in the response matrix.

The LOWARP application

One nice feature with PLS is that all the model diagnostic tools which we have considered so far, that is, R^2/Q^2, ANOVA, residual plots, and so on, are retained. But, in addition to these diagnostic tools, PLS provides other model parameters, which are useful for model evaluation and interpretation. We will now describe some of these PLS parameters and illustrate how they are used in the interpretation of a PLS regression model.

Factors

	Name	Abbr.	Units	Type	Use	Settings
1	glass	gl	Fraction	Formulation	Controlled	0.2 to 0.4
2	crtp	cr	Fraction	Formulation	Controlled	0 to 0.2
3	mica	mi	Fraction	Formulation	Controlled	0 to 0.2
4	amtp	am	Fraction	Formulation	Controlled	0.4 to 0.6

Responses

	Name	Abbr.	Units	Transform	Prec.	MLR Scale	PLS Scale
1	wrp1	w1		None	Free	None	Unit Variance
2	wrp2	w2		None	Free	None	Unit Variance
3	wrp3	w3		None	Free	None	Unit Variance
4	wrp4	w4		None	Free	None	Unit Variance
5	wrp5	w5		None	Free	None	Unit Variance
6	wrp6	w6		None	Free	None	Unit Variance
7	st1	st1		None	Free	None	Unit Variance
8	st2	st2		None	Free	None	Unit Variance
9	wrp7	w7		None	Free	None	Unit Variance
10	st3	st3		None	Free	None	Unit Variance
11	st4	st4		None	Free	None	Unit Variance
12	wrp8	w8		None	Free	None	Unit Variance
13	st5	st5		None	Free	None	Unit Variance
14	st6	st6		None	Free	None	Unit Variance

Figure 24.1: *(left) Factors varied in the LOWARP application.*
Figure 24.2: *(right) Measured responses in the LOWARP application.*

The application that we will present concerns the production of a polymer with certain desired properties. Four factors, ingredients, were varied according to a 17 run mixture design, that is, each run of the design corresponded to one polymer. These factors are presented in Figure 24.1. To map the properties of the polymers produced, 14 responses were measured. Eight responses reflecting the warp and shrinkage of the polymers, and six responses expressing the strength of the polymers, were measured. These responses are summarized in Figure 24.2. The desired combination was low warp/shrinkage and high strength. The worksheet used is listed in Figure 24.3. As seen, small amounts of missing data are found among the responses, but this is not a problem when employing PLS.

ExpNo	ExpName	RunOrder	InOut	glass	crtp	mica	amtp	wrp1	wrp2	wrp3	wrp4	wrp5	wrp6	st1	st2	wrp7	st3	st4	wrp8	st5	st6
1	N1	16	In	0.4	0.1	0.1	0.4	0.9	5	0.2	1	0.3	4.2	232	15120	1.2	2190	26390	1.3	2400	0.7
2	N2	1	In	0.2	0.2	0	0.6	3.7	7.3	0.7	1.8	2.5	5.4	150	12230	1.8	905	20270	2.1	1020	0.6
3	N3	11	In	0.4	0.2	0	0.4	3.6	6.9	0.9	2.1	4.8	9.4	243	15550	1.2	1740	21180	1.4	1640	
4	N4	6	In	0.2	0.2	0.2	0.4	0.6	3.1	0.3	0.4	0.4	1.1	188	11080	1	1700	17630	1	1860	0.5
5	N5	13	In	0.2	0.1	0.2	0.5	0.3	2.1	0.3	0.3	0.8	1.1	172	11960	1.2	1810	21070	1.3	1970	0.5
6	N6	3	In	0.4	0	0.2	0.4	1.2	5					245	15600	1.1	2590	25310	1.3	2490	0.6
7	N7	4	In	0.2	0	0.2	0.6	2.3	3.9	0.3	0.4	0.7	1.4	242	13900	1.5	1890	21370	1.6	1780	
8	N8	7	In	0.4	0	0.1	0.5	2.6	5.9	0.4	0.2	0.7	1.2	243	17290	1.6	2130	30530	1.6	2320	0.7
9	N9	14	In	0.3	0.2	0.1	0.4	2.2	5.3	0.2	0.7	0.6	2	204	11170	1	1670	19070	1.1	1890	0.6
10	N10	5	In	0.4	0	0	0.6	5.8	7	0.9	1	5.6	11.8	262	20160	1.6	1930	29830	1.8	1890	
11	N11	12	In	0.3	0	0.2	0.5	0.8	2.9	0.5	0.6	1.1	2	225	14140	1.3	2140	22850	1.3	2110	0.7
12	N12	15	In	0.3	0.1	0	0.6	2.8	5.1	1	1.2	2.7	6.1	184	15170	1.9	1230	23400	2.1	1250	0.6
13	N13	9	In	0.3	0.1	0.1	0.5	1.1	4.7	0.6	0.9	1.3	3.5	198	13420	1.4	1750	23790	1.4	1930	0.7
14	N14	8	In	0.3	0.1	0.1	0.5	1.9	4.7	1	1	2.8	5.4	234	16970	1.5	1920	25010	1.6	1790	0.7
15	N15	10	In	0.3	0.1	0.1	0.5	2.9	5.9	0.5	0.6	1	6.6	239	15480	1.5	1800	23140	1.6	1730	
16	N16	2	In	0.4	0.1	0	0.5	5.5	7.9	0.8	2.4	5.5	9.3	256	18870	1.5	1880	28440	1.8	1790	
17	N17	17	In	0.3	0	0.1	0.6	3.2	6	0.3	0.5	1.5	5.2	249	16310	1.5	1860	24710	1.7	1780	

Figure 24.3: *The LOWARP experimental data.*

PLS model interpretation – Scores

In the analysis of the LOWARP data it is recommended to start by evaluating the raw experimental data. We made a histogram and a replicate plot for each one of the 14 responses, but could detect no anomalies in the data. No plots of this phase are provided. We will now describe the results of the PLS analysis. Conceptually, PLS may be understood as computing pairs of new variables, known as *latent variables* or *scores*, which summarize the variation in the responses and the factors. When modelling the LOWARP data, three such pairs of latent variables, three *PLS-components*, were obtained. In Figure 24.4, we can see how R^2 and Q^2 evolve as a result of calculating each consecutive pair of latent variables.
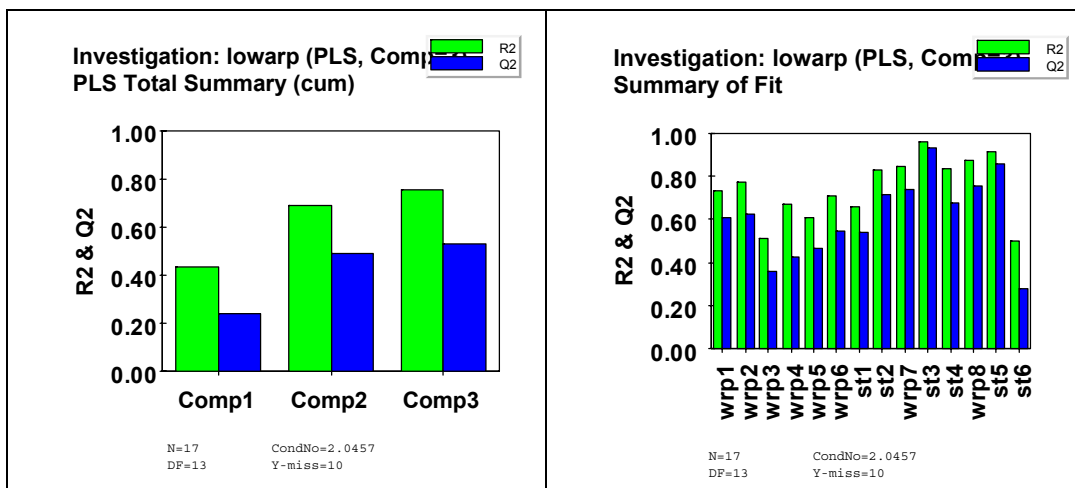
Figure 24.4: *(left) Summary of fit per PLS component.*
Figure 24.5: *(right) Summary of fit per response variable, after 3 components.*

After three components $R^2 = 0.75$ and $Q^2 = 0.53$. These are excellent values considering the fact that 14 responses are modelled at the same time. It is possible to fractionate these overall performance statistics into R^2's and Q^2's related to the individual responses. This is shown in Figure 24.5. The individual R^2-values range between $0.50 – 0.96$ and Q^2-values between $0.28 – 0.93$, but the interesting feature is that for no response are the individual R^2/Q^2-values separated by more than 0.25. This constitutes a strong indication that the multivariate model is well-founded and warranted.

Furthermore, the three pairs of latent variables offer the possibility to visually inspect the correlation structure between the factors and the responses. Figure 24.6 provides a scatter plot of the first pair of latent variables. Here, the first linear summary of Y, called u_1, is plotted against the first linear summary of X, called t_1, and each plot mark depicts one polymer, one experiment in the design. With a perfect match between factors (X) and responses (Y), all samples would line up on the diagonal going from the lower left-hand corner to the upper right-hand corner. Conversely, the weaker the correlation structure between factors and responses, the more scattered the samples are expected to be around this ideal diagonal. Any point remote from this diagonal may be an outlier. We can see in Figures 24.6 and 24.7 that for the two first PLS-components, there is a strong correlation structure. The correlation band for the third PLS-component, as shown in Figure 24.8, is not as distinct, but still reasonably strong, considering it is a third model component. In

conclusion, these plots indicate that the factor changes are influential for the monitored responses, and that there are no outliers in the data. This is vital information.
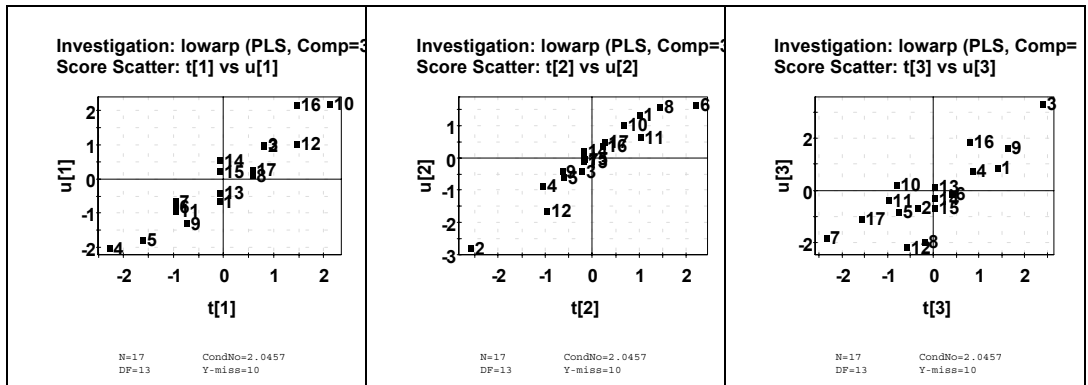

Figure 24.6: (left) PLS score plot of first model dimension.
Figure 24.7: (middle) PLS score plot of second model dimension.
Figure 24.8: (right) PLS score plot of third model dimension.

PLS model interpretation – Loadings

Because PLS results in only one model, it is possible to overview the relationships between *all* factors and *all* responses at the same time. An efficient means of interpreting the PLS model is the loading plot displayed in Figure 24.9. We have plotted the loadings of the second model dimension, denoted wc_2, against the loadings of the first model dimension, denoted wc_1.
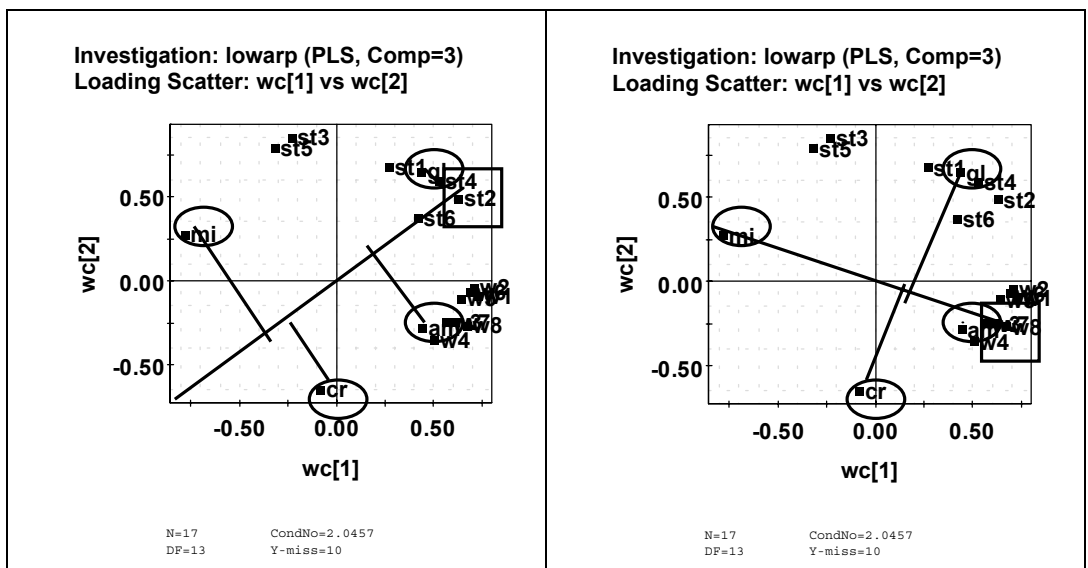
Figure 24.9: (left) PLS loadings – interpretation of st2.
Figure 24.10: (right) PLS loadings – interpretation of w8.

In the model interpretation one considers the distance to the plot origin. The further away from the plot origin an X- or Y-variable lies, the stronger the model impact that particular variable has. In addition, we must also consider the sign of the PLS loading, which informs about the correlation among the variables. For instance, the X-variable (factor) glas is influential for the Y-variable (response) st4. This is inferred from their closeness in the loading plot. Hence, when glas increases st4 increases.

Figure 24.9 contains some extra lines which are meant to support the interpretation process. First, the interpreter has to focus on an interesting response, for example the boxed st2. From the location of this response one then draws a reference base-line through the origin. Subsequently, orthogonal projections are carried out from the position of each factor and onto this first reference line. The resulting intercepts will help us in the interpretation; the further away from the plot origin the more potent the factor. Hence, we understand that glas and mica are most important for st2, albeit with different relationships to the response. The latter factor is negatively correlated with st2, since the intercept with the reference line is located on the opposite side of the plot origin.

Now, if we want to interpret the model with regard to another response, say w8, we have to draw another reference base-line and make new orthogonal factor projections. This is displayed in Figure 24.10. For the w8 response, mica is the most important factor, and exerts a negative influence, that is, to decrease w8 we should increase mica. The factors glas and amtp are positively correlated with w8, but their impact is weaker than mica. The last factor, crtp, is the least influential in the first two model dimensions. This is inferred from the fact that the projection intercept is almost at the plot origin.

PLS model interpretation – Coefficients

As mentioned before, it is possible to transfer the PLS solution into an expression based on regression coefficients. We will now demonstrate the practical part of this, not the matrix algebra involved. Consider Figure 24.11, in which four responses of the LOWARP application have been marked.
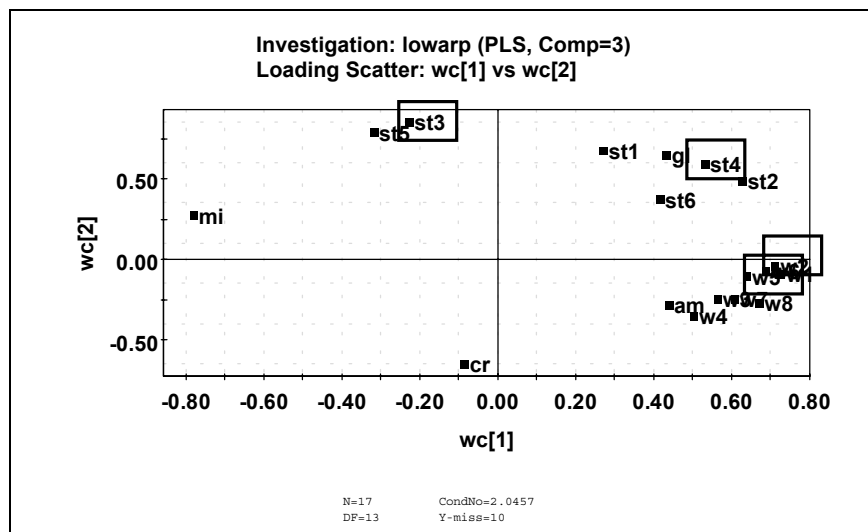
Figure 24.11: The PLS loadings of the two first model dimensions. The four responses w2, w6, st3, and st4 are marked for comparative purposes (see text).

The responses denoted w2 and w6 lie almost on top of each other and their labels are hard to discern. These represent a pair of strongly correlated responses. Based on their orientation in the loading plot, we expect them to have similar regression coefficient profiles. That this is indeed the case is evidenced by Figures 24.12 and 24.13.
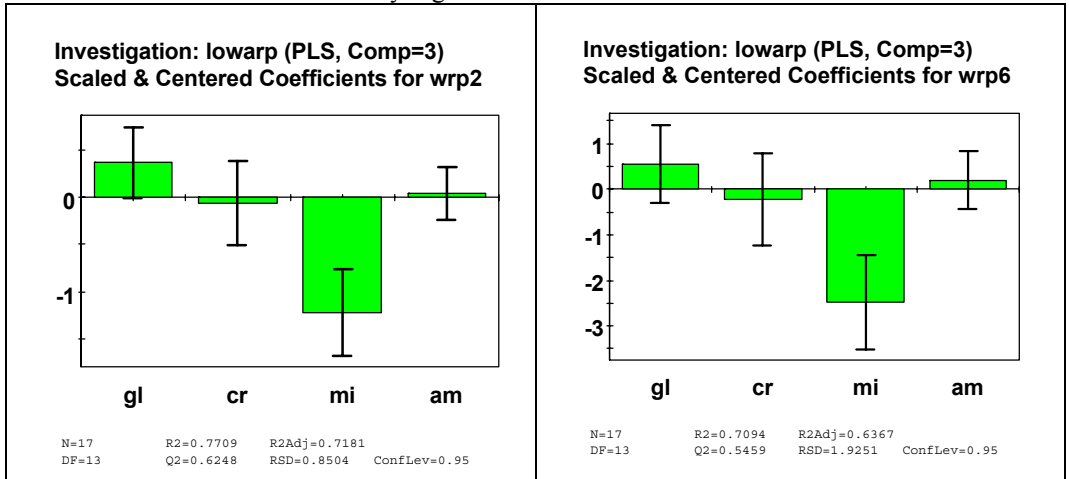
Figure 24.12: (left) Regression coefficients for w2.
Figure 24.13: (right) Regression coefficients for w6.

The other variable pair is made up of st3 and st4. Since these are located in different regions of the loading plot, they should not have similar regression coefficient profiles. Figures 24.14 and 24.15 indicate that these profiles mainly differ with regard to the third and fourth factors.
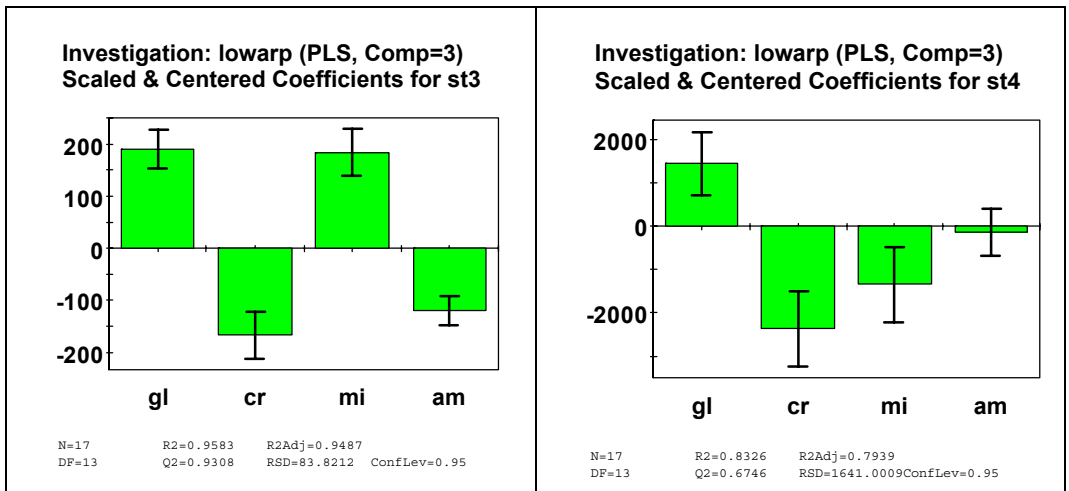
Figure 24.14: (left) Regression coefficients for st3.
Figure 24.15: (right) Regression coefficients for st4.

This ability of the loading plot to overview the relationships among the responses is very useful in the model interpretation. It will facilitate the understanding of which responses provide similar information about the experimental runs, and which provide unique

information. The implication of a loading plot like the one in Figure 24.11 is that it is not necessary to measure 14 responses in the future. It is sufficient to select a sub-set of responses representing a good spread in the loading plot. Both categories of responses are grouped in two clusters. Hence, a proper selection might be st1 and st2 to represent the st-grouping in the upper right-hand quadrant, st3, w1-w2 for one w-group, and w3-w4 for the other w-group. This corresponds to a reduction by 50%.

PLS model interpretation – VIP

So far, we have only considered the PLS loadings associated with the two first model components. However, in order to conduct the model interpretation stringently, we ought to consider the loadings of the third component as well. With three PLS loading vectors, three bivariate scatter plots of loadings are conceivable, and these are plotted in Figures 24.16 – 24.18.
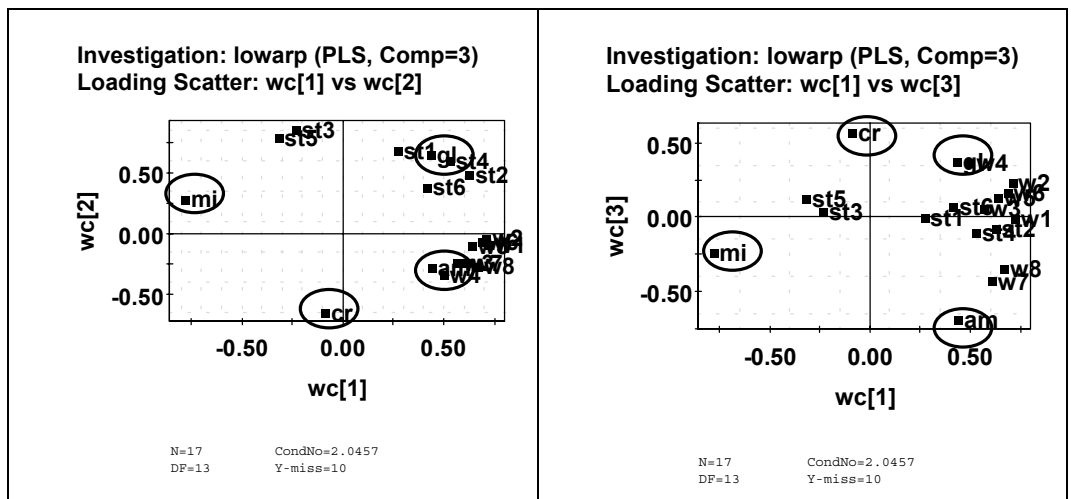
Figure 24.16: (left) Scatter plot with loading vector 2 versus loading vector 1.
Figure 24.17: (right) Scatter plot with loading vector 3 versus loading vector 1.

From a model overview perspective, it is quite natural to primarily consider the loading plot displayed in Figure 24.16, as this plot represents most of the modelled response variation. However, if details are sought, one must also consider the loadings of the higher-order model dimensions. With many PLS components, say four or five, the interpretation of loadings in terms of scatter plots becomes impractical. Also, with as many as 14 responses, the model interpretation based on regression coefficients is cumbersome, since 14 plots of coefficients must be created. In this context of many PLS components and many modelled responses, PLS provides another parameter which is useful for model interpretation, the VIP parameter (see Figure 24.19).
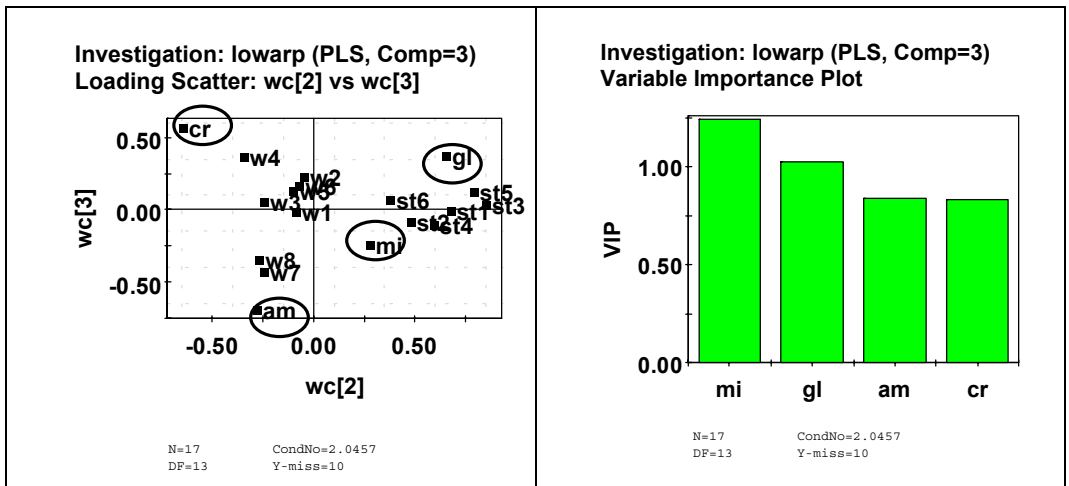
Figure 24.18: *(left) Scatter plot with loading vector 3 versus loading vector 2.*
Figure 24.19: *(right) VIP plot.*

VIP is the acronym for *variable importance in the projection*. This parameter represents the most condensed way of expressing the PLS model information. VIP is a weighted summary of all loadings and across all responses, and hence there can be only one VIP-expression per PLS model. Model terms with high VIP-values, often higher than $0.7 – 0.8$, are the most influential in the model. We can see in Figure 24.19 that mica and glas are the two most important factors. In the case of many model terms, many more than four, the VIP plot often displays discrete jumps, like stairs, and it is possible to define a suitable threshold below which factors might be discarded from the model. Therefore, in complicated PLS applications, a plot of the VIP-parameter is often a useful start in the model interpretation for revealing which are the dominating terms. Such information is subsequently further explored by making appropriate plots of loadings or coefficients.

The linear PLS model – Matrix and geometric representation

Now, that we have seen how PLS operates from a practical point of view, it is appropriate to try to understand how the method works technically. The development of a PLS model can be described as follows: For a certain set of observations – experimental runs in DOE – appropriate response variables are monitored. These form the N x M response data matrix Y, where N and M are the number of runs and responses, respectively. This is shown in Figure 24.20. Moreover, for the same set of runs, relevant predictor variables are gathered to constitute the N x K factor matrix X, where N is the same as above and K the number of factors. The Y-data are then modelled by the X-data using PLS.
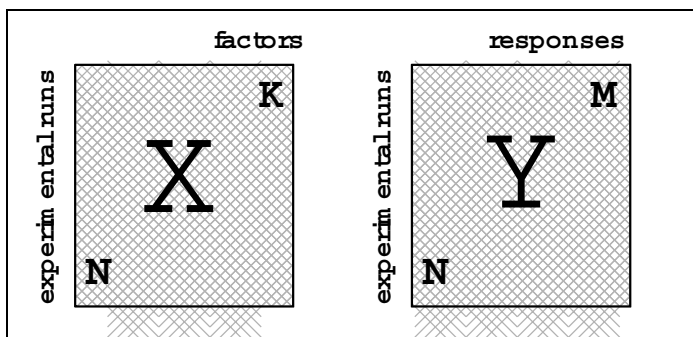
Figure 24.20: *Some notation used in PLS.*

A geometric representation of PLS is given in Figure 24.21. The experimental runs can be seen as points in two spaces, that of X with K dimensions and that of Y with M dimensions. PLS finds lines, planes or hyperplanes in X and Y that map the shapes of the point-swarms as closely as possible. The drawing in Figure 24.21 is made such that PLS has found only one-dimensional models in the X- and Y-spaces.
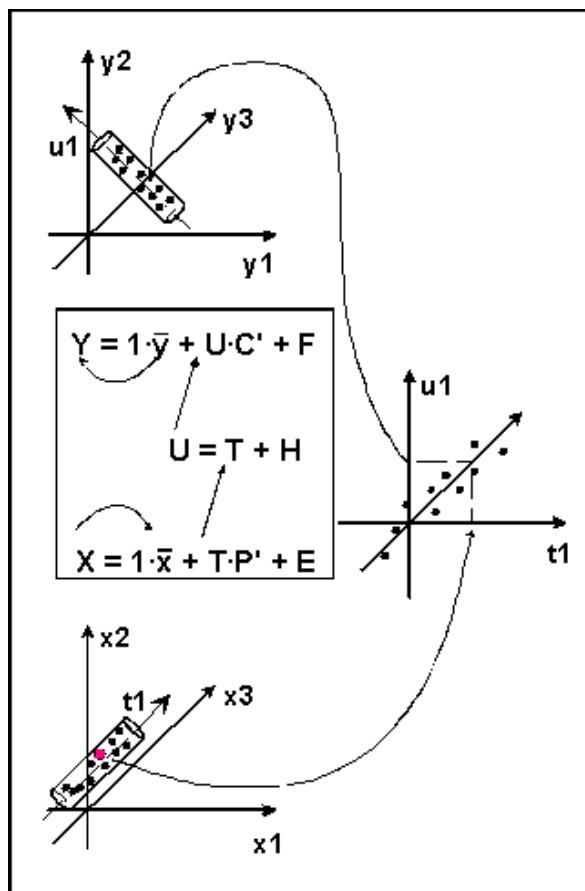

Figure 24.21: *Geometric representation of PLS.*

PLS has two primary objectives, namely to well approximate X and Y and to model the relationship between X and Y. This is accomplished by making the bilinear projections

X = TP' + E   (eqn 24.1)

Y = UC' + G   (eqn 24.2)

and connecting X and Y through the inner relation

U = T + H   (eqn 24.3)

where E, G and H are residual matrices. A more detailed account of the PLS algorithm is given in the statistical appendix. PLS simultaneously projects the X- and Y-variables onto the same subspace, T, in such a manner that there is a good relation between the position of one observation on the X-plane and its corresponding position on the Y-plane. Moreover, this relation is asymmetric (X ➔ Y), which follows from equation (3). In this respect, PLS differs from for instance canonical correlation where this relation is symmetric.

The linear PLS model – Overview of algorithm

In essence, each PLS model dimension consists of the X score vector t, the Y score vector u, the X loading vector p', the X weight vector w' and the Y weight vector c'. This is illustrated in Figure 24.22, which provides a matrix representation of PLS. Here, the arrows show the order of calculations within one round of the iteration procedure. Solid parts of the arrows indicate data "participating" in the computations and dashed portions "inactive" data. For clarity, the X score matrix T, the X loading matrix P', the X weight matrix W', the Y score matrix U, and the Y weight matrix C' are represented as matrices, though in each iteration round only the last vector is updated.
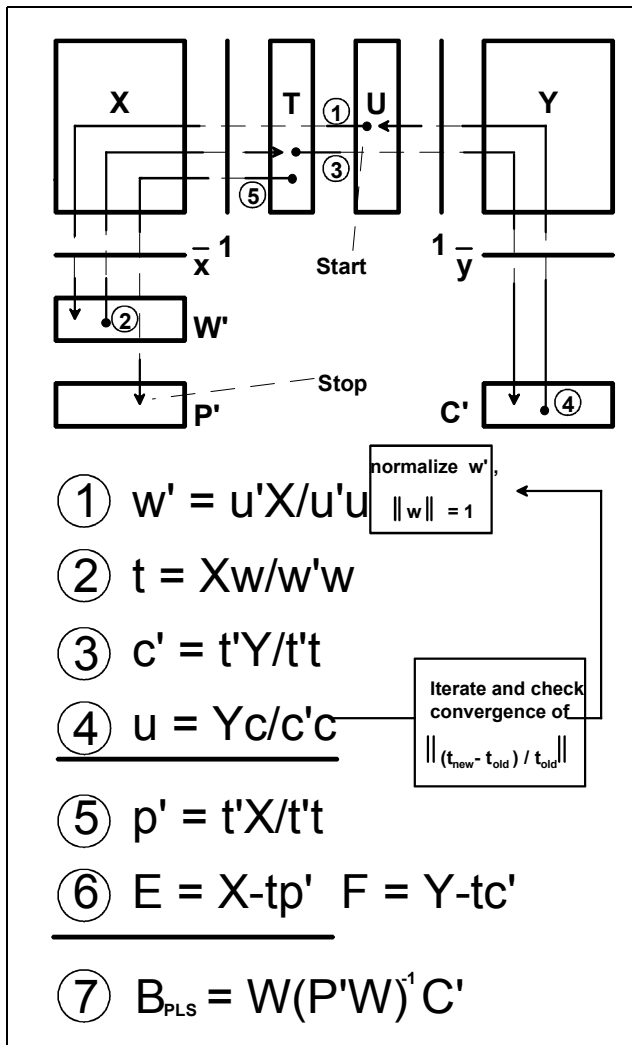
Figure 24.22: *An overview of the PLS algorithm.*

It is the PLS vectors w' and c', formally known as *weights* but in everyday language known as *loadings*, that are used in the interpretation of each PLS component. Based on this argumentation one may see that another way to understand PLS is that it forms "new x-variables", t, as linear combinations of the old ones, and thereafter uses these new t's as predictors of Y. Only as many new t's are formed as are needed, and this is assessed from their predictive power using a technique called cross-validation. Cross-validation is described in the statistical appendix.

Summary of PLS

Once a PLS model has been derived, it is important to interpret its meaning. For this, the scores t and u can be considered. They contain information about the experimental runs and their similarities / dissimilarities in X- and Y-space with respect to the given problem and

model. To understand which factors and responses are well described by the model, PLS provides three complementary interpretational parameters. These are the loadings, the coefficients, and VIP, which are exemplified in Figures 24.23 – 24.25.
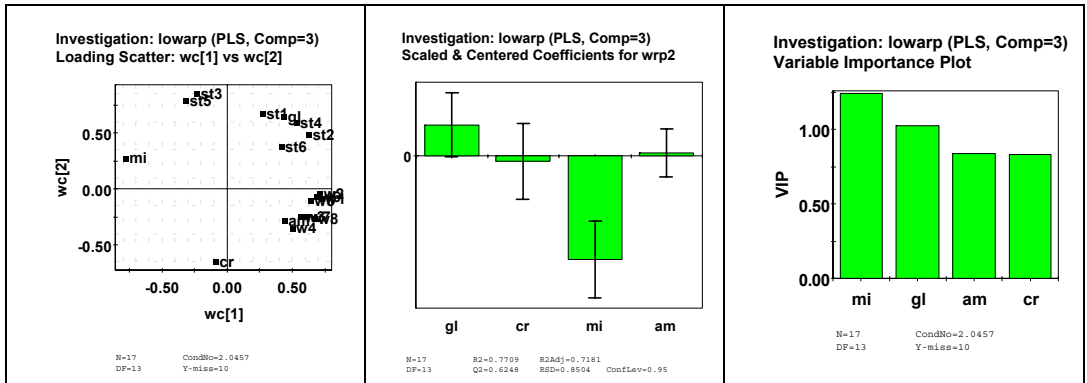
Figure 24.23: *(left) PLS parameters for model interpretation – loadings.*
Figure 24.24: *(middle) PLS parameters for model interpretation – coefficients.*
Figure 24.25: *(right) PLS parameters for model interpretation – VIP.*

The loadings wc provide information about how the variables combine to form t and u, which, in turn, expresses the quantitative relation between X and Y. Hence, these loadings are essential for the understanding of which X-variables are important for modelling Y (numerically large w-values), for understanding of which X-variables provide common information (similar profiles of w-values), and for the interpretation of the scores t.

Sometimes it may be quite taxing to overview the PLS loadings and in such circumstances the VIP parameter is useful. VIP in square is a weighted sum of squares of the PLS loadings, w, taking into account the amount of Y-variance explained. Alternatively, the PLS solution may be transferred into a regression-like model:

$$Y = X \, B_{PLS} + F \qquad \text{(Eqn 24.4)}$$

Here B_{PLS} corresponds to the regression coefficients. Thus, these coefficients are determined from the underlying PLS model and can be used for interpretation, in the same way as coefficients originating from MLR. However, with collinear variables, which occur for instance in mixture design applications, we must remember that these coefficients are *not* independent. Further, the parts of the data that are not explained by the model, the residuals, are of diagnostic interest. Large Y-residuals indicate that the model is inadequate, and a normal probability plot of the residuals of a single Y-variable is useful for identifying outliers.

Quiz

Please answer the following questions:

When is it appropriate to use PLS?

Which diagnostic tools are available through PLS?

What is a latent variable?

Summary

PLS is a multivariate regression method which is useful for handling complex DOE problems. This method is particularly useful when (i) there are several correlated responses in the data set, (ii) the experimental design has a high condition number, or (iii) there are small amounts of missing data in the response matrix. We used the LOWARP application with 14 responses and some missing Y-data to exemplify the tractable properties of PLS. Interestingly, PLS contains the multiple linear regression solution as a special case, when one response is modelled and the condition number is unity. Model interpretation and evaluation is carried out analogously to MLR, the difference being that with PLS more informative model parameters are available. Particularly, we think here of the PLS loadings, which inform about the correlation structure among all factors and responses, and the PLS scores, which are useful for finding trends, groups, and outliers.

Statistical Appendix

Designs

Classical designs used in MODDE are described in the chapter: Objective, Modeling and Design. The default generators used by MODDE for fractional factorial designs are those recommended by Box, Hunter and Hunter (page 410). You may edit and change the generators by using the Edit generator menu. When you update the confounding, MODDE will warn you if some of the effects in your model are confounded with each other, i.e. if your model is singular.

For more information see Box, G.E.P., Hunter, W.G., and Hunter, J.S., (1978) "Statistics for Experimenters" New York, Wiley.

Analysis

MODDE supports Multiple Linear Regression (MLR) and PLS for fitting the model to the data.

Model

You may edit the model and add or delete terms. You may add up to third order terms (cubic terms, or 3 factor interaction, etc.).

If your design is singular with respect to your model, MODDE will fit the model with PLS, and MLR will not be available.

Hierarchy

MODDE enforces hierarchy of the model terms. You cannot delete the constant term. You can only delete a linear term if no higher order term containing the factor is still in the model.

Scaling X

When fitting the model with multiple regression, the design matrix X is scaled and centered as specified in the factor definition box: MLR scaling. If the choice is not orthogonal, the condition number will differ from the one displayed using Menu Analysis: Worksheet: Evaluate.

When fitting the model with PLS the X matrix is always scaled to unit variance.

If warranted, the scaled X matrix is extended with squares and / or cross terms according to the selected model.

The choices of scaling are:

(x denotes the original factor value and z the scaled one)

Orthogonal scaling: $z_i = \dfrac{(x_i - M)}{R}$

Where M = midrange, R = Range/2.

Midrange scaling: $z_i = (x_i - M)$

Unit variance scaling: $z_i = \dfrac{(x_i - m)}{s}$

Where m = average, s = standard deviation (computed from WS).

Note that Orthogonal and Midrange scaling are only available with MLR.

MODDE default scaling for MLR is the orthogonal scaling.

Scaling Y

The matrix of responses Y, when fitting the model with PLS is scaled by default, to unit variance. You can modify the unit variance scaling weight by using the PLS scaling box in the factor definition. With MLR the Y's are not scaled.

Evaluate

Computes the condition number of the orthogonally scaled and centered extended design matrix using the SVD (Singular value decomposition). The X matrix is taken from the Worksheet. The calculation depends on fit method (MLR, PLS) and which factors are involved. A message informs the user how the condition number is calculated.

Condition Number

The condition number is the ratio of the largest and the smallest singular values of X (eigenvalues of X'X). This condition number represents a measure of the sphericity of the design (orthogonality). All factorial designs, without center-points have a condition number of 1 and the design points are situated on the surface of a sphere.

Missing Data

Missing data in X

Missing data in X are not allowed, and will disable the fit. This also applies to uncontrolled X-variables. If the user stills want to analyze the data, an arbitrary value can be "filled in" at the place of the missing value, and the row then deleted from the calculations by setting it as "out" in the 4th Worksheet column ("In/Out").

Missing data in Y

Multiple Regression

All rows with missing data in any Y are excluded from the analysis for all Y's, hence N, displayed on plots and lists is the number of observations without missing data in any Y.

PLS

With PLS, missing data are handled differently. When all Y-values are missing in a row, that row is excluded from the analysis. When there are some "present" Y-data in a row, the row is NOT excluded, but included in the projection estimation in PLS. This leads, however, to minor differences in the displayed N and DF at the bottom of plots and lists.

ANOVA, R2adj, and Residual Plots

The N-value used in ANOVA, and for the computation of R^2adj, is the actual number of non-missing observations for each response-column. This N-value and DF = N-p are displayed at the bottom of the ANOVA plots and lists, the, and on all residual plots, including observed vs. predicted Y.

Confidence Intervals

Confidence intervals of coefficients, and predictions, are computed using the total number of observations, regardless of missing values. This total number of observations is displayed as N at the bottom of all other plots and lists. This approximation is possible because the confidence intervals computed with the regression formulas are somewhat too large because the PLS solution is a shrunk estimator with smaller prediction errors than those of regression. Hence a small number of missing elements in Y does not make the PLS confidence intervals larger than those computed with the regression formulas and the total number of observations.

Residual Standard Deviation

The residual standard deviation displayed in the summary table and at the bottom of all plots and lists including ANOVA is computed with the total number of observations without excluding the missing values. This is the RSD used in the computation of confidence interval, for coefficients and predictions.

Y-miss

At the bottom of all plots and lists, we display Y-miss = the number of missing elements in the total Y matrix (all responses). Y-miss is always equal to 0, when fitting with MLR or with PLS and there are no missing data.

Multiple Linear Regression (MLR)

Multiple regression is extensively described in the literature, and this chapter will only identify the numerical algorithms used to compute the regression results, the measures of goodness of fit and diagnostics used by MODDE. For additional information on MLR, see Draper and Smith "Applied Regression Analysis", Second Edition, Wiley, New York.

MODDE uses the singular value decomposition (SVD) to solve the system of equations:

$$Y = X*B + E$$

Y is an n*m matrix of responses, and X (the extended design matrix) an n*p matrix, with p the number of terms in the model including the constant, and B the matrix of regression coefficients, E the matrix of residuals. See Golub and Van Loan (1983) for a description of the SVD and its use to obtain the regression results.

In case of missing data in a row (x or/and y), this row is eliminated before the MLR fitting.

Partial Least Squares (PLS)

PLS is briefly described at the end of this chapter. MODDE calculates the PLS model using the PLS2 NIPALS algorithm.

PLS tolerates a small amount of missing data in y (maximum 10%).

The following results are computed for both MLR and PLS.

ANOVA

The analysis of variance (ANOVA) partitions the total variation of a selected response SS (Sum of Squares corrected for the mean) into a part due to the regression model and a part due to the residuals.

$$SS = SS_{regr} + SS_{resid}$$

If there are replicated observations, the residuals sum of squares is further partitioned into pure error SS_{pe} and Lack of fit SS_{lof}.

$$SS_{resid} = SS_{lof} + SS_{pe}$$

$$DF_{resid} = (n - p)$$

$$SS_{pe} = \sum_{ki} (e_{ki} - e_k)^2$$

$$DF_{pe} = \sum_{k} (n_k - 1)$$

$$DF_{lof} = n - p - \sum_{k} (n_k - 1)$$

n = number of experimental runs (exc. missing values)

nk = number of replicates in the kth set

p = number of terms in the model, including the constant

ek = average of the nk residuals in the kth set of replicates

j = jth residual in the kth set of replicates

A goodness of fit test is performed by comparing the MS (mean square) of lack of fit to the MS of pure error:

$$F_{lof} = \frac{(SS_{lof} / DF_{lof})}{(SS_{pe} / DF_{pe})}$$

Two ANOVA plots are displayed:

- The Regression goodness of fit test
- The LOF goodness of fit test

Checking for Replicates

MODDE checks the rows of the Worksheet for replicates. Rows in the Worksheet are considered replicates if they match all factor values plus or minus a 5% tolerance.

Measures of Goodness of Fit

MODDE computes and displays the following:

Q^2

$$Q^2 = \frac{(SS - PRESS)}{SS}$$

With

$$PRESS = \sum_i \frac{(Y_i - \hat{Y}_i)^2}{(1 - h_i)^2}$$

h_i is the ith diagonal element of the Hat matrix

$$X(X'X)^{-1}X'$$

R^2

$$R^2 = \frac{(SS - SS_{resid})}{SS}$$

$$R^2_{adj} = \frac{(MS - MS_{resid})}{MS}$$

$$MS = SS / (n-1), \quad MS_{resid} = SS_{resid} / (n-p)$$

RSD = Residual standard deviation = $\sqrt{MSE_{resid}}$

Degrees of Freedom

MODDE always computes the real degrees of freedom RDF of the residuals:

$$RDF = n - p - \sum_k (n_k - 1)$$

n = number of experimental runs

n_k = number of replicates in the kth set

p = number of terms in the model, including the constant

Saturated Models

When RDF = 0 the model is saturated, and MODDE does not compute or display R^2, R^2 Adjusted or Q^2 when fitting the model with MLR. With PLS only Q^2 is computed and displayed.

Singular models

Singular models (condition number > 3000) are only fitted with PLS.

If $p > n - \sum_k (n_k - 1)$, the degrees of freedom of the residuals are computed as:

DF_{resid} = 0, if no replicates in the design

DF_{resid} = $\Sigma_k(n_k - 1)$, if replicates in the design

Coefficients

The regression coefficients computed and displayed by MODDE refer to the centered and scaled data. You may also select to display the "unscaled and uncentered" coefficients.

Normalized Coefficients

In the overview plot, to make the coefficients comparable between responses, the "centered and scaled" coefficients are normalized with respect to the variation of Y, that is they are divided by the standard deviation of their respective Y's.

PLS Orthogonal Coefficients

The "centered and scaled" coefficients of PLS refer to factor values scaled to unit variance.

The PLS orthogonal coefficients re-express the coefficients to correspond to factors centered and orthogonally scaled, i.e. using the Midrange and Low and High values from the Factor Definition.

Confidence Intervals

For matrices with condition number < 3000, MLR and PLS computes confidence intervals on coefficients as:

$$\sqrt{(X'X)^{-1}} * RSD * t(\alpha/2, DF_{resid})$$

For matrices with condition number > 3000, PLS does not compute confidence intervals on the coefficients.

Coding qualitative variables at more than 2 levels

If a term in the model comprises a qualitative factor, C, at k levels, there will be k-1 expanded terms associated with that term. For example, if the levels of the qualitative factor C are (a, b, c, and d) the three expanded terms C (j) are as follows:

C	C(2)	C(3)	C(4)
a	-1	-1	-1
b	1	0	0
c	0	1	0
d	0	0	1

The coefficients of these expanded terms are given as the coefficients for level 2 (b), 3 (c), and 4 (d) of C, while the coefficient for level 1 (a) is computed as the negative sum of the three others. MODDE displays all the four coefficients in the coefficient table but notes that they are associated with only three degrees of freedom.

Residuals

Raw Residuals

The raw residual is the difference between the observed and the fitted (predicted) value

$$e_i = Y_i - \hat{Y}_i$$

The raw residuals are displayed in the residual lists.

Standardized Residuals

The standardized residual is the raw residual divided by the residual standard deviation

$$e_i \: / \: s \:\: (s = RSD)$$

These are MODDE default for PLS Residual plots.

Deleted Studentized Residuals

With MLR, for models with 2 or more degrees of freedom, deleted studentized residuals is MODDE default when plotting residuals.

Deleted studentized residuals are not available with PLS.

The deleted studentized residual is the raw residual e_i divided by its "deleted" standard deviation (s_i) which is the residual standard deviation (s_i) computed with observation (i) left out of the analysis, and corrected for leverage, i.e.:

$$e_i = \frac{e_i}{s_i \sqrt{(1 - h_i)}}$$

S_i = is an estimate of the residual standard deviation with observation i left out of the model

h_i is the ith diagonal element of the Hat matrix: $X(X'X)^{-1}X'$

For more information see Belsley, Kuh, and Welsch (1980).

Deleted studentized residuals requires at least two degrees of freedom.

Predictions

For X matrices with condition number < 3000, both MLR and PLS computes a confidence interval for the average predicted y:

$$Y_i + \sqrt{h_i} * RSD * t(\alpha / 2, DF_{resid})$$

h_i is the ith diagonal element of the Hat matrix: $X(X'X)^{-1}X'$

For X matrices with condition number > 3000 and for all Cox mixture models, PLS computes only the standard error of the average predicted Y:

$$SE(Y) = \sqrt{X_0' * (TT)^{-1} * X_0} * RSD$$

Box-Cox Plot

A useful family of transformation on the necessarily positive Y's is given by the power transformation:

$$Z = Y^\lambda \qquad \text{for } \lambda \text{ not equal to zero}$$

$$Z = \ln Y \qquad \text{for } \lambda \text{ equal zero}$$

MODDE, for values of λ between -2 and 2, computes L_{max} and plots it against the values of λ with a 95% confidence interval.

$$L_{max}(\lambda) = \frac{1}{2}\ln(SS_{resid} / n) + (\lambda - 1) * \sum \ln Y_i$$

SS_{resid} is the Residual Sum of Squares after fitting the model $Z = X * \beta + e$ for the selected value of λ.

The value of λ that maximizes $L_{max}(\lambda)$ is the maximum likelihood estimator of λ.

For more information see Draper and Smith "Applied Regression Analysis, Second Edition" Wiley, New York.

The Box-Cox plot is not available for PLS.

Partial Least Squares (PLS)

When several responses have been measured, it is useful to fit a model simultaneously representing the variation of all the responses to the variation of the factors. PLS deals with many responses simultaneously, taking their covariance's into account. This provides an overview of the relationship between the responses and of how all the factors affect all the responses. This multivariate method of estimating the models for all the responses simultaneously is called PLS.

PLS contains the multiple regression solution as a special case, i.e. with one response and a certain number of PLS dimensions, the PLS regression coefficients are identical to those obtained by multiple regression.

PLS has been extensively described in the literature and only a brief description is given here. PLS finds the relationship between a matrix Y (response variables) and a matrix X (predictor or factor variables) expressed as:

$$Y = XB + E$$

The matrix Y refers to the characteristics of interest (responses). The matrix X refers to the predictor variables and to their squares or/and cross terms if these have been added to the model.

PLS creates new variables (t_a) called X-scores as weighted combinations of the original X-variables: $t_a = Xw_a$, where w_a are the combinations weights. These X-scores are few, often just two or three, and orthogonal (independent). The X-scores are then used to model the responses.

With several responses, the Y-variables are similarly combined to a few Y-scores (u_a) using weights c_a, $u_a = Yc_a$. The PLS estimations is done in such a way that it maximizes the correlation, in each model dimension, between t_a and u_a, One PLS component (number a) consists of one vector of X-scores (t_a), and one of Y-scores (u_a), together with the X and Y-weights (w_a and c_a).

Hence the PLS Model consists of a simultaneous projection of both the X and Y spaces on a low dimensional hyper plane with new coordinates T (summarizing X) and U (summarizing Y), and then relating U to T. This analysis has the following two objectives:

To well approximate the X and Y spaces by the hyperplanes

To maximize the correlation between X and Y (between u and t).

Mathematically the PLS model can be expressed as:

$$X = TP' + E$$
$$Y = TC' + F$$

Geometrically, we can see the matrices X and Y as n points in two spaces, (see figure), the X-space with p axes, and the Y-space with m axes, p and m being the number of columns in X (terms in the model) and in Y (responses).
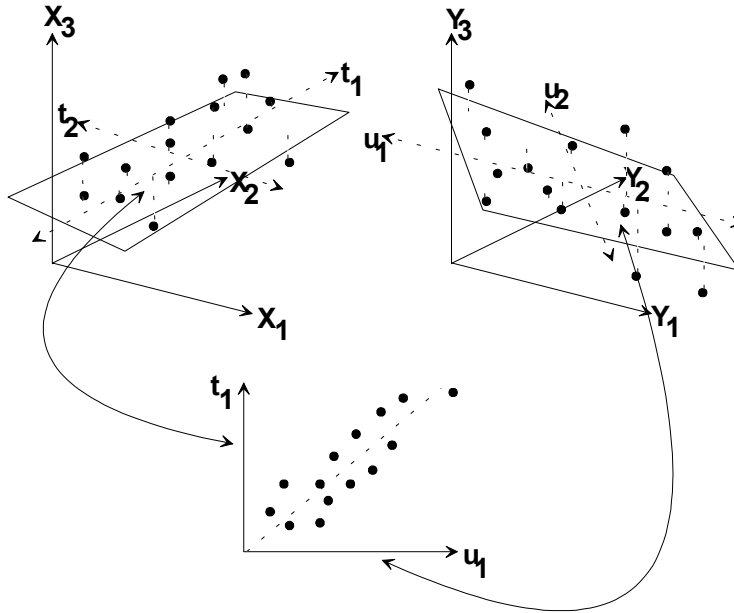
The model dimensionally, (number of significant PLS components), is determined by cross validation (CV), where PRESS (see below) is computed for each model dimension. One selects the number of PLS dimensions that give the smallest PRESS.

Model predictive power

The predictive power of an MLR or a PLS model is given by Q^2 which is based on the Prediction Residual Sum of Squares, PRESS. This is a measure of how well the model will predict the responses for new experimental condition. The computations are repeated several times with, each time, different objects kept out of the calculation of the model. PRESS is then computed as the squared differences between observed Y and predicted values Ypred when the objects (rows in the tables X and Y) were kept out from the model estimation. Q^2 is computed as:

$$Q^2 = \frac{(SS - PRESS)}{SS}$$

here SS = sum of squares of Y corrected for the mean. A Q^2 larger than zero indicates that the dimension is significant (predictive). An overall Q^2 is computed for all PLS components, for all the responses and for each individual response, and represent the percent variation of Y that is predictive. Large Q^2, 0.7 or larger, indicates that the model has good predictive ability and will have small predictions errors. Q^2 is the predictive measure corresponding to the measure of fit, R^2, (the percent variation of the response explained by the model).

$$R^2 = \frac{(SS - SS_{resid})}{SS}$$

Q^2 gives a lower estimate to how well the model predicts the outcome of new experiments, while R^2 gives an upper estimate.

Automatic Cross Validation Rules:

A PLS component is cross validated if:

Rule 1: PRESS for all Y's together < 1.2

or

Rule 2: PRESS for at least $M^{1/2}$ Y's < 1.2

and

Rule 3: SS explained for all Y' s together > 1%

or

Rule 4: SS explained for all separate Y's > 2%

MODDE computes a minimum of two PLS components (if they exist), even if not significant.

PLS Plots

Both scores and loading plots are available:

Loading Plots

WC plots (PLS)

Plots of the X- and Y-weights (w and c) of one PLS dimension against another, say, no.'s 1 and 2, show how the X-variables influence the Y-variables, and the correlation structure between X's and Y's. In particular one better understands how the responses vary, their relation to each other and which ones provide similar information.

Score Plots

TT, UU, and TU plots (PLS)

The tt and uu plots, of the X- and Y-scores of, say, dimensions 1 and 2 (i.e. t_1 vs. t_2, and u_1 vs. u_2), can be interpreted as windows into the X- and Y-spaces, respectively, showing how the design points (experimental conditions, X) and responses profile (Y) are situated with respect to each other. These plots show the possible presence of outliers, groups, and other patterns in the data.

The tu plots (t_1 vs. u_1, t_2 vs. u_2, etc.) show the relation between X and Y, and display the degree of fit (good fit corresponds to small scatter around the straight line), indications of curvature, and outliers.

PLS Coefficients

PLS computes regression coefficients (B_m) for each response Y_m expressed as a function of the X's according to the assumed model (i.e. linear, linear plus interactions or quadratic,). These coefficients are (columns of B) computed as:

$$B = W(P'W)^{-1}C'$$

W and C are (p*A) and (m*A) matrices whose columns are the vectors w_a and c_a.

p = number of terms in the model

m = number of responses

A = Number of PLS components

Mixture Experiments

In a Mixture experiment the responses of interest depend only on the relative proportions of the components (called mixture factors) that make up the mixture or formulation. Hence, the sum of all the mixture factors is a constant **T**, usually equal to 1 when no mixture factors are kept constant.

Mixture and Process Factors

Mixture factors are expressed as the fraction of the total amount of the formulation. Their experimental ranges lie between 0 and 1.

Regular factors (i.e., temp, pH, etc.) that are not part of the mixture or formulation are referred to as process factors. These are expressed as amounts or levels, and can be either quantitative (measured on a continuous scale) or qualitative (have only discrete values).

MODDE supports both mixture and process factors in the same experiment.

Mixture Factors

A mixture factor can be a formulation factor or a filler factor. Only one mixture factor can be defined as filler.

Formulation

These are the usual mixture factors used in formulations with specifically defined experimental ranges. Most Mixture experiments have only formulation factors.

Filler

The presence of a filler is typical of certain types of simple mixture experiments. For example in a synthesis the solvent is a typical filler, as is water in a juice punch. A filler is a mixture component, usually of little interest, making up a large percentage of the mixture, and added at the end of a formulation to bring the mixture total to the desired amount.

It is recommend to define a Mixture factor as a **filler** when:

It is always present in the mixture, **and**

It accounts for a large percentage of the mixture there is no restriction on its range. It is added at the end to bring up the mixture total to the desired amount (usually 1 when no mixture factors are kept constant) **and**

You are not interested in the effect of the filler per se.

Note: *Only one mixture factor can be defined as filler.*

When you specify a filler factor, MODDE checks that these conditions are met and defaults to a slack variable model, with the filler factor omitted from the model.

Use

All mixture factors are controlled or constant. The uncontrolled button is grayed out for both formulation and filler factors.

Formulation factors can be defined as "Constant" when you want to keep them constant in the experiment.

When mixture factors are constant, the mixture total $T = 1$ - Sum (constant mixture factors). When no formulation factors are defined as constant, the mixture total has to be equal to 1. MODDE issues an error message and stops whenever the mixture total is not equal to T or 1.

Note: *A filler factor cannot be "Constant".*

Scaling

Mixture factors are always unscaled when you fit the model with MLR. When you fit the model with PLS, all mixture factors are scaled to unit variance.

Note: *When the mixture region is regular, mixture factors are first transformed to pseudo components, and then scaled with PLS models. It is also selectable from the Options menu.*

Mixture Constraint

In a mixture experiment the mixture total (i.e. the sum of all the mixture factors in the experiment) is equal to a constant T. The mixture Total T is generally equal to 1 when no mixture factor is kept constant. This mixture constraint implies that the mixture factors are not independent, and this collinearity has implications on the mixture experimental region, the mixture designs, and the mixture model formulation.

Mixture Designs

Mixture Experimental Region

When all mixture factors vary from 0 to T (the mixture total), the shape of the experimental region is a **Simplex**. With constraints on their ranges, the experimental region is usually an irregular polyhedron inside the simplex. In some constrained cases, as for example, with lower bounds constraints only, the experimental region is a small simplex inside the original simplex. See Crosier (1984).

MODDE checks for consistent bounds, and computes:

$R_U = \Sigma U_i$ - T

$R_L = T - \Sigma L_i$

L_i and U_i are the lower and upper bound of the **ith** mixture factors.

From R_L, R_U and R_i (the range of every formulation factor) MODDE determines if the experimental region is a Simplex (the L simplex oriented as the original one, or the U simplex with opposite orientation) or an irregular polyhedron.

Regular region

Pseudo Components transformations

When the mixture region is the L or U simplex, MODDE defaults to transforming the mixture factors to pseudo component to make all their ranges vary between 0 and 1. This is very similar to orthogonal scaling of process factor, to make their ranges vary between -1 and +1.

With regular mixture region, MODDE uses classical mixture designs.

The design is expressed in pseudo components and the worksheet is of course always displayed in original units.

The analysis is performed on the mixture factors transformed to pseudo component, as the coefficients of the Cox model can then be directly interpreted as the mixture factors effects.

Note: *You can select in the option menu to have the analysis done on the mixture factors expressed in original units.*

Region is the L Simplex

When the mixture region is the L simplex, the L pseudo component transformation is defined as:

$P_i = (X_i - L_i)/(R_L)$

The transformed mixture factors P_i vary from 0 to 1.

The Region is U Simplex

When the mixture region is the U simplex, the U pseudo component transformation is defined as:

$P_i = (U_i - X_i)/(R_U)$

The transformed mixture factors P_i vary from 0 to 1, but in this case the new simplex in the P's has an opposite orientation to the original simplex in X, that implies that effects in P are reversed from those in X.

Classical Mixture Designs

When all factors are mixture factors and the shape of the region is a Simplex, the designs available in MODDE are the following classical mixture designs (all classical mixture designs are displayed in "Show design" in Pseudo components, and by default the analysis is done with the formulation factors transformed to pseudo components).

Screening Design

MODDE provides three variant of the axial design. Axial designs locates all the experimental points on the axis of the simplex and are recommended for screening, see Snee (references).

Standard Axial (AXN)

This design includes the following $2*q + m$ runs (q = number of mixture factors, m centroid points as specified by user).

1. All the q vertex points. The coordinate of the **i**[th] Vertex point is
   $x_i = (0, 0, 0..1, 0, 0..)$.

---

2. All q interior points of the simplex. The coordinate of the ith Interior point is $x_i = (1/2q, 1/2q, 1/2q,..(q+1)/2q, 1/2q, 1/2q..)$.

3. The overall centroid of the simplex with coordinate
   $x = (1/q, 1/q,....., 1/q..)$ replicated (m-1) times.

Extended Axial (AXE)

This design includes the following $3*q +m$ runs (q = number of mixture factors, m specified by user).

1. All the q vertex points. The coordinate of the ith Vertex point is $x_i = (0, 0, 0..1, 0, 0..)$.

2. All q interior points of the simplex. The coordinate of the ith Interior point is $x_i = (1/2q, 1/2q, 1/2q,..(q+1)/2q, 1/2q, 1/2q..)$.

3. All the q End points. The coordinate of the ith End point is
   $x_i = (1/(q-1), 1/(q-1), 1/(q-1), 0, 1/(q-1), 1/(q-1)..)$.

4. The overall centroid of the simplex with coordinate
   $x = (1/q, 1/q,....., 1/q..)$ replicated (m-1) times.

Reduced Axial (AXR)

The AxR includes the following (q+m) (specified by user) points:

1. All the q vertex points.

2. A subset or none (specified by user) selected from the **q** interior points.

3. The Overall centroid replicated as desired.

RSM

Quadratic Models

MODDE provides 2 variants of the Simplex Centroid. The Simplex centroid design has all the experimental points on the vertices, and on the center of the faces of consecutive dimensions.

Modified simplex Centroid (SimM)

This design includes the following:

1. The q vertex points. The coordinate of the ith Vertex point is
   $x_i = (0, 0, 0..1, 0, 0..)$.

2. The $(q (q-1))/2$ Edge centers. The coordinate of the ij^{th} edge point is $x_{ij} = (0, 0, 1/2, 1/2 0, 0..)$.

3. The q Interior check points. The coordinate of the ith interior point is $x_i = (1/2q, 1/2q, (q+1)/2q, 1/2q, 1/2q..)$.

4. The overall centroid with coordinate $x = (1/q, 1/q,...1/q)$, replicated as desired.

Modified simplex Centroid Face Center (SimF)

This design includes the following:

1. The q vertex points. The coordinate of the ith Vertex point is
   $x_i = (0, 0, 0..1, 0, 0..)$.

2. The $(q (q-1))/2$ Edge centers. The coordinate of the ij^{th} edge point is $x_{ij} = (0, 0, 1/2, 1/2 0, 0..)$.

3. The q Face centers of dimension (q-1). The coordinate of the **i**th face center is: (1/q-1, 1/q-1,..,0, 1/q-1..1/q-1).

4. The q Interior check points. The coordinate of the **i**th interior point is x_i = (1/2q, 1/2q, (q+1)/2q, 1/2q, 1/2q..).

5. The overall centroid with coordinate x = (1/q, 1/q,...1/q), replicated as desired.

Special Cubic Model

The Simplex Centroid Special Cubic (SimSC)

This design includes the following:

1. The q vertex points. The coordinate of the **i**th Vertex point is
   x_i = (0, 0, 0..1, 0, 0..).

2. The (q (q-1)) 1/3, 2/3 Edge points. The coordinate of the **ij**th edge point is
   x_{ij} = (0, 0, 1/3, 2/3, 0, 0..), x_{ji} = (0, 0, 2/3, 1/3, 0, 0..).

3. The q(q-1)(q-2)/6 Face centers of dimension 2. The coordinate of the **i**th face center is
   (0, 0, 0, 1/3, 1/3, 1/3..0, 0,).

4. The q Interior check points. The coordinate of the **i**th interior point is x_i = (1/2q, 1/2q, (q+1)/2q, 1/2q, 1/2q..).

5. The overall centroid with coordinate x = (1/q, 1/q,...1/q), replicated as desired.

Cubic Model

The Simplex Centroid Cubic (SimC)

This design includes the following:

1. The q vertex points. The coordinate of the **i**th Vertex point is
   x_i = (0, 0, 0..1, 0, 0..).

2. The (q (q-1)) 1/3, 2/3 Edge points. The coordinate of the **ij**th edge point is
   x_{ij} = (0, 0, 1/3, 2/3, 0, 0..), x_{ji} = (0, 0, 2/3, 1/3, 0, 0..).

3. The q(q-1)/2 Edge centers. The coordinate of the **i**th edge center is x_i = (0, 0, 0, 1/2, 1/2, 0...0).

4. The q(q-1)(q-2)/6 Face centers of dimension 2. The coordinate of the **i**th face center is
   (0, 0, 0, 1/3, 1/3, 1/3..0, 0,).

5. The q Interior check points. The coordinate of the **i**th interior point is x_i = (1/2q, 1/2q, (q+1)/2q, 1/2q, 1/2q..).

6. The overall centroid with coordinate x = (1/q, 1/q,...1/q), replicated as desired.

Irregular Region: D-Optimal Designs

Screening

When the mixture region is an irregular polyhedron, MODDE computes the extreme vertices (corners) delimiting the region. These extreme vertices constitute the candidate set and center of the high dimensional faces are added to support potential terms. (see D-Optimal designs). The design is a D-optimal selection of N (specified by user) runs from the candidate set.

RSM

MODDE computes the extreme vertices, 1/3, 2/3 centers of edges, centers of faces of dimension (q-1) and the overall centroid of the experimental region. When there are too many extreme vertices, only the center of the 25% longest edges are computed. These experimental points constitute the candidate set.

The design is a D-Optimal selection of N runs (specified by the user) from the Candidate set.

Pseudo Component Transformation

You can always select to have the mixture factors expressed in pseudo components for the analysis. MODDE uses the L pseudo component transformation when $R_L \leq R_U$ and the U pseudo component when $R_U < R_L$.

Pseudo component transformation is the MODDE default when the method of fit is MLR as it stretches the experimental region and alleviates the problem of ill conditioning.

Note: *All mixture designs are displayed in pseudo components.*

Mixture Models

Because of the mixture constraint, (the mixture factors are not independent) the analysis of mixture data with multiple regression requires a special model form.

The traditional approaches have been:

- Defining the model omitting one mixture factor, hence making the others independent. This is the Slack Variable approach.

- Omitting some terms from the model, so that the terms remaining in the model are independent. This is Scheffé model, with the constant term removed from the linear model and the quadratic terms removed from the quadratic model.

- Using the complete model including all the mixture terms, but putting constraints on the coefficients to make them estimable. This is the Cox reference model, and the constraints on the coefficients are defined with respect to a standard reference mixture. This standard reference mixture serves the same function as the centering constant with process variables models.

Analysis of mixture data in MODDE

Mixture factors only

Model forms

Slack Variable Model

When you define a mixture factor as filler, MODDE generates the slack variable model by omitting the filler factor from the model. The model is generated according to the selected objective and is treated as a non mixture model. You may select MLR or PLS to fit the model as with ordinary process factors. With MLR the factors will be orthogonally scaled.

Cox Reference Model

When all mixture factors are formulation factors, MODDE generates, by default, the Cox reference model, i.e. the complete polynomial model linear or quadratic. MODDE supports also a special cubic and a full cubic model (see chapter on Objective and models).

Scheffé Model

You may select to fit a Scheffé model, by selecting MLR Scheffé model, in the Option Menu. MODDE expresses the mixture model in the Scheffé form. The full cubic model is not supported as a Scheffé model.

Analysis and Fit Method

In MODDE, the default "Fit method" with mixture factors is PLS, and the model form is the Cox reference mixture model. All factors, including mixture factors are scaled to unit variance, by default, prior to fitting. This is also done with mixture factors that have been transformed to pseudo components.

The Cox Reference Model

The Cox reference model can be fitted by MLR (when obeying mixture hierarchy) and in all cases by PLS

The coefficients in the Cox model are meaningful and easy to interpret. They represent the change in the response when going from a standard reference mixture (with coordinates s_k) to the vertex k of the simplex. In other words when component x_k changes by Δ_k, the change in the response is proportional to b_k. Terms of second or higher degree are interpreted as with regular process variables models. The presence of square terms, though they are not independent, facilitates the interpretation of quadratic behavior, or departure from non linear blending. The constant term is the value of the response at the standard reference mixture.

- **Changing the Standard Reference Mixture.**
  Use "Edit: Reference Mixture" menu item to change the coordinates sk of the standard reference mixture. By default MODDE selects as reference mixture the centroid of the constrained region.

- **Mixture Hierarchy with the Cox reference model.**
  By default all Cox reference models, linear and quadratic, obey "mixture hierarchy". That is the group of terms constrained by:
  $\Sigma b_k s_k = 0$
  $\Sigma c_{kj} b_{kj} s_k = 0$ for k = 1,,,,q (1) and for j = 1,,,,q
  ($c_{kj} = 1$ when $j \neq k$ and $c_{kj} = 2$ when k = j.)
  are treated as a unit, and terms cannot be removed individually.

If you want to remove terms individually, as with regular process models, turn off the "Mixture Hierarchy" in the "Edit menu". When the "Mixture Hierarchy" is off,(this includes cubic models), the Cox reference model can only be fitted by PLS. The coefficients are the regular PLS coefficients computed from the projection and not re-expressed relative to a stated standard reference mixture. Note that in all cases model hierarchy is enforced (a term cannot be removed, if a higher order term containing that factor is still in the model).

ANOVA with the Cox Model

In the "Analysis of Variance" table, the degrees of freedom for regression are the real degrees of freedom, taking into account the mixture constraint. These are the same as the equivalent slack variable model.

Screening Plots

When the objective is to find the component effects on the response, the coefficients of the Cox reference linear model are directly proportional to the Cox effects. The Cox effect is the change in the response when component k varies from 0 to 1 along the Cox axis. That is the axis joining the reference point to the **k**[th] vertex.

Effect Plot

This plot displays the adjusted Cox effects. The adjusted effect of component **k** is equal $r_k * t_k$ r_k range of factor k equal

$r_k = U_k - L_k$ and tk is the total Cox effect, defined as:

$t_k = b_k/(T - s_k)$, T = mixture total, and b_k the unscaled coefficients.

This plot is only available for screening designs using the Cox model.

Main Effect Plot

For a selected mixture factor X_k, this plot displays the predicted change in the response when X_k varies from its **low** to its **high** level, adjusted for all other mixture factors, that is, by default, the relative amounts of all other mixture factors are kept in the same proportion as in the standard reference mixture (MODDE does not check if the other mixture factors are kept within their ranges).

For example, if the Main effect of the mixture factor X_1 is being displayed, when X_1 takes the value x1, the other mixture factors are assigned the values: $x_j = (T - x_1) * (s_j / T - s_1)$.

S_k are the coordinates of the standard reference mixture. The standard reference mixture is the one used in the model.

You can change this default and select to have all other mixture factors kept in the same proportion as their ranges (this ensures no extrapolation).

Interaction Plot

Interaction plot are not available when you only have mixture factors.

MLR and the Cox Reference model

In MODDE you can fit Cox reference mixture models (linear or quadratic) with MLR, only when they obey mixture hierarchy. When fitting the model with MLR the mixture factors are not scaled and are only transformed to pseudo components when the region is regular. The model is fitted by imposing the following constraints on the coefficients:

Linear models

$\Sigma b_k s_k = 0$ (1)

Quadratic models

$\Sigma b_k s_k = 0$ (1)

$\Sigma c_{kj} b_{kj} s_k = 0$ for k = 1,,,,q (1) and for j = 1,,,,q (2)

Here $c_{kj} = 1$ when $j \neq k$ and $c_{kj} = 2$ when k = j.

Design of Experiments - Principles and Applications

and s_k are the coordinates of the standard reference mixture.

PLS and the Cox reference mixture

With PLS the standard reference mixture is not stated a priori as with multiple regression, and no constraints on the coefficients are explicitly imposed. PLS fits the mixture models, and deals with all collinearities by projecting on a lower dimensional subspace. The PLS coefficients can be interpreted as in the Cox model, relative to a reference mixture resulting from the projection, but not explicitly stated.

Expressing PLS coefficients relative to an explicit standard Reference Mixture

With linear and quadratic models obeying hierarchy,, it is easy to re-express the PLS coefficients relative to a stated reference mixture with coordinates s_k. (s_k expressed in pseudo component, if pseudo component transformation was used)

On the fitted PLS model one imposes the following constraints, on the uncentered, unscaled coefficients (See Cox).

Linear models

$\Sigma b_k s_k = 0$ (1)

Quadratic models

$\Sigma b_k s_k = 0$ (1)

$\Sigma c_{kj} b_{kj} s_k = 0$ for k = 1,,,,q (1) and for j = 1,,,,q (2)

Here $c_{kj} = 1$ when $j \neq k$ and $c_{kj} = 2$ when $k = j$.

The scaled and centered coefficients are recomputed afterwards.

Note: *In MODDE, with linear and quadratic models obeying the mixture hierarchy, (i.e. terms constrained by (1) or (2) can only be removed as a group, and not individually), by default the PLS coefficients are always expressed relative to a stated standard reference mixture.*

With models containing terms of the third order (cubic), or disobeying mixture hierarchy, no constraints are imposed on the PLS coefficients. The coefficients are in this case, the regular PLS coefficients and the reference mixture is implicit and results from the projection.

Scheffé Models derived from the Cox model

With the linear or quadratic Cox reference model, one can re-express the unscaled coefficients as those of a Scheffé model. The following relationship holds:

Linear

Scheffé, $b_k = $ Cox (PLS) $b_0 + b_k$

Quadratic

Scheffé, $b_k = $ Cox (PLS) $b_0 + b_k + b_{kk}$

Scheffé, $b_{kj} = $ Cox (PLS) $b_{kj} - b_{kk} - b_{jj}$

MLR Solution derived from PLS

Because PLS contains multiple regression as a special case, you can with PLS derive the same solution as when you fit the Cox model with Multiple Regression. When you extract as many PLS components as available you get the same solution as MLR.

With MODDE you do the following:

- First Fit the model. MODDE extracts only the significant PLS components. This is the PLS solution.

- Then use the "Next Component" menu and continue extracting PLS components until no correlation between X and Y remains. This is the MLR solution.

Scheffé Models

The Scheffé models are only fitted with MLR and only the main effect plot is available.

Scheffé models are only available for investigation with all mixture factors.

ANOVA

As described by Snee in "Test Statistics for Mixture Models" (Technometrics, Nov. 1974), the degrees of freedom in the ANOVA table are computed in the same way as with the slack variable model.

Using the Model

Y-Predicted Plot

This plot is available for all objectives and all model forms. As with process factors, this plot displays a spline representing the variation of the fitted function, when the selected mixture factor varies over its range, adjusted for the other factors. As with the main effect plot, this means that the relative amounts of all other mixture factors are kept in the same proportion as in the standard reference mixture. If no standard reference mixture is specified, the centroid of the constrained region is used as default.

Contour Plots: Mixture Contour

Trilinear contour plots are available with mixture factors and but no 3D plots.

Process and Mixture Factors

When you have both process and mixture factors, you can select to treat them as one model, or to specify separate models for the mixture factors, and the process factors.

With both mixture and process factors, the only model form available is the Cox reference mixture model.

When the model obeys mixture hierarchy, the PLS coefficients are expressed relative to a stated standard reference mixture. The following constraints are imposed on the coefficients:

For linear models

$\Sigma b_k s_k = 0$

For Quadratic models

$\Sigma b_k s_k = 0$ (1)

$\Sigma c_{kjbkj}s_k = 0$ for $k = 1,,,,q$ (1) and for $j = 1,,,,q$ (2)

Here $c_{kj} = 1$ when $j \neq k$ and $c_{kj} = 2$ when $k = j$.

If γ (gamma) are the coefficients of the interactions between the process and mixture factors:

$\Sigma \gamma_k s_k = 0$

Note: *When the model contains terms of order 3, or contains qualitative and formulation factors, the PLS coefficients are not adjusted relative to a stated standard mixture.*

MODDE Plots

All of MODDE plots are available when you have both mixture and process factors. For both the "Main Effect" and "Y-Predicted" plots, when you select to vary a process factor, all of the mixture factors are set to the values of the standard reference mixture. When you select to vary a mixture factor, process factors are set on their average and the other mixture factors are kept in the same proportion as in the standard reference mixture or their ranges.

Inclusions

Inclusions are extra runs that will be part of the worksheet.

Classical Designs

With all classical designs, Inclusions are simply added to the worksheet after the design and worksheet are generated. They are not included in the "Desired number of runs" specified by the user. It makes no difference if the Inclusions are specified before or after the worksheet is generated.

If the worksheet already exists when you enter the inclusions, click on the "Add to Worksheet" button. MODDE adds the inclusions to the end of the worksheet and issues a message to confirm.

If you enter the inclusions before the generation of the worksheet, click on OK to close this window. After you generate the worksheet, get back to the "Inclusions" menu and click on the "Add to Worksheet" button.

Note: The inclusions are added to the worksheet only when you click on the "Add to Worksheet" button.

You can delete any or all inclusions rows from the worksheet, be by using the right mouse button "Delete" command.

D-Optimal Designs

With D-Optimal designs Inclusions runs can either be part of the design or added at the end of the worksheet.

Inclusions as part of the design

When you specifies inclusions before the design and worksheet are generated, mark the inclusion box in "generate D-optimal" and the inclusions will be used as runs in the D-Optimal design.

Note: *The number of runs N specified in the objective include the number of inclusions.*

Inclusions added to worksheet

If you want the inclusions to be added at the end of the worksheet, you can:

- Enter the inclusions after the worksheet has been generated, and click on the "Add to Worksheet" button.

- Enter the inclusions prior to the worksheet generation, and do not mark the box "Use inclusions" in the generate D-Optimal menu. After you generate the worksheet, use the "Inclusions" menu to click on "Add to Worksheet".

Note: *When you add the inclusions at the end of the worksheet, they are not included in the number of runs N, specified in the objective.*

If he says no, then he has to go to "generate D-Optimal" and regenerate the D-Optimal design and then the inclusion will be part the D-Optimal design and included in the selected number of runs

These runs cannot be deleted.

Note: *When generating D-Optimal designs if inclusions exists (they have been specified) they are always part of the design and included in the number of runs.*

D-Optimal Designs

What are D-Optimal designs

D-Optimal designs are computer generated designs, tailor made for your problem. They allow great flexibility in the specifications of your problem. They are particularly useful when you want to constrain the region and no classical design exist.

"D-Optimal" means that these designs maximize the information in the selected set of experimental runs with respect to a stated model.

For a specified regression model $y = X*\beta + \varepsilon$ where:

y is a (N x 1) vector of observed responses, X is a (N x p) extended design matrix, i.e. the n experimental runs extended with additional columns to correspond to the p terms of the model (i.e., the added columns are for the constant term, interaction terms, square terms, etc..)

beta is a (p x 1) vector of unknown coefficients to be determined by fitting the model to the observed responses.

epsilon is a (N x 1) vector of residuals (the differences between the observed and predicted values of the response y). They are assumed to be independent of each other, normally distributed and with constant variance σ^2

The D-Optimal design maximizes the determinant of the X'X matrix, which is an overall measure of the information in X. Geometrically, this corresponds to maximizing the volume of X in a p dimensional space.

When to use D-Optimal designs

Whenever possible you should use classical designs and these are the default designs of MODDE. However when classical designs are impossible to apply, D-Optimal designs is the preferred choice.

MODDE suggests a "D-Optimal" design in the following cases:

1. You have linear constraints on the factor settings, reducing the experimental region to an irregular polyhedron.

2. There are no classical designs that can well investigate the irregular region. D-Optimal design is then the preferred choice as it makes efficient use of the entire experimental space.

3. You have formulation factors, with lower and upper bounds, and possibly additional constraints, making the region an irregular polyhedron.

4. You have specified qualitative factors, with more than two levels and there are no mixed level design available.

5. Your objective is RSM and you have qualitative factors.

6. The number of experimental runs you can afford is smaller than the number of runs of any available classical design.

Candidate Set

D-Optimal designs are constructed by selecting N runs from a candidate set. This candidate set is the discrete set of "all potential good runs".

MODDE generates the candidate set as follows:

I) For a regular process region, the candidate set consists of one or more of the following sets of points (depending on your model and the number of factors):

• The full factorial for up to 10 factors, reduced factorial for up to 32 factors.

• Centers of edges between hyper-cube corners

• Centers of the faces of the hyper-cube.

• Overall centroid

II) For constrained regions of mixture or/and process factors, the candidate set consists of one or more of the following set of points:

The extreme vertices of the constrained region

The centers of the edges. If these exceed 200, the center of the 200 longest edges

The centers of the various high dimensional faces

The overall centroid.

MODDE has implemented an algorithm to compute the extreme vertices, center of edges, center of faces etc. as described by Piepel (1988).

The D-Optimal Algorithm

D-Optimal designs have been criticized for being too dependent on an assumed model. To reduce the dependence on an assumed model, MODDE has implemented a Bayesian Modification of the K-Exchange algorithm of Johnson and Nachtsheim (1983), as described by W. DuMouchel and B. Jones in "A Simple Bayesian Modification of D-Optimal designs to reduce dependence on an Assumed Model", Technometrics (1994).

With this algorithm one can add to the "primary terms" i.e. the terms in the model, "potential terms", i.e. additional terms that might be important. The objective is to select a

---

D-Optimal design, rich enough to guard for potential terms, and enable the analysis to detect possibly active ones.

In order not to increase the number of runs N, and to avoid a singular estimation, one assumes that the coefficients of the potential terms are likely to have a mean of 0 and a finite variance (tau, τ)^2.

Implementation of the D-Optimal algorithm in MODDE

K-exchange algorithm

The k-exchange algorithm is a compromise between the exchange algorithm of Wynn (1972) with k=1 and the Federov algorithm with k = N (the selected number of runs). **In MODDE k is set to 3**, that is at every iteration of the procedure, the algorithm considers an exchange between k = 3 points in the design with the smallest prediction variance and points in the candidate set. If any exchange increases the determinant, the point(s) (up to 3) are exchanged.

Variance of the coefficients of the potential terms

As recommended by W. DuMouchel. tau, τ, is set to 1 in MODDE.

Potential terms

In this version of MODDE the potential terms are added automatically and cannot be modified by the user. In future versions the user will be able to alter this default selection of potential terms.

Note: *You can remove potential terms all together.*

Depending on the number of factors, the objective and the model, MODDE adds the following potential terms:

Process Factors with constraints

Screening

Factors	Model	Potential terms
2-12	Linear	All Interactions
2-12	Linear + Interactions	All squares

RSM

Factors	Model	Potential terms
2-8	Quadratic	All cubes

Process Factors without Constraints

Screening

Factors	Model	Potential terms
2-20	Linear	All interactions
21-32	Linear	Interactions between

		the first 20 factors
2-17	All Interactions	Squares

RSM

Factors	Model	Potential terms
2-6	quadratic	all cubes
7-12	quadratic	None

Note: *No potential terms are added for investigations with all factors defined as qualitative.*

Mixture Factors and irregular regions

Screening

Factors	Model	Potential terms
2-20	linear	all squares + interactions

RSM

Factors	Model	Potential terms
2-12	quadratic	All cubes

Number of D-Optimal Designs

You can select to have several "D-Optimal designs" generated at once by the "Generate D-Optimal" menu.

Design Evaluation

To Evaluate and compare D-Optimal designs, MODDE computes the following criteria:

LogDetNorm

The log of the determinant of X'X normalized for number of terms in the model p, and number of runs N.

This is the criterion used, by default, to select the best design. MODDE selects the design with the largest value (closest to 0) of Logdetnorm.

$$\text{Logdetnorm} = Log_{10}\left[\frac{(X'X)^{1/p}}{N}\right]$$

The maximum value of Logdetnorm, for an orthogonal design, is 0.

LogDet

The Log of the determinant of the X'X matrix

Condition No

The condition number of the X design matrix, coded orthogonally, and extended according to the model.

Geff

G efficiency is a lower bound on D efficiency, which compares the efficiency of a D-Optimal design to a Fractional Factorial.

G efficiency is defined as: $\text{Geff} = \dfrac{100*p}{n*d}$

where

p = number of terms in the model

n = number of runs in the design

d = Maximum relative prediction variance v over the candidate set, the prediction variance $v = x(X'X)^{-1}x'$

Inclusions and Design Augmentation

MODDE allows you to specify a set of experimental runs as "Inclusions". These runs, can be part of the resulting "D-Optimal Design".

Inclusions are useful for Design Augmentation. If you already have a worksheet with N experimental runs, and say, want to add M additional experiments, (to de-confound interactions, or estimate curvature, etc.), specify the worksheet as "inclusions", ask for N+M runs and state the desired model. The M runs are selected D-Optimally from the candidate set with respect to your model.

Optimizer

The optimizer uses a Nelder Mead simplex method with the fitted responses functions, to optimize an overall desirability function combining the individual desirability of each response.

Desirability

For every response y, the desirability function is computed as follows:

$f(g(y)) = 100*(e^{\lambda g(y)} - 1)$

with $g(y) = 100*((y - P)/(T - P))$

T, L and P are defined as follows:

T = user desired Target

L = User defined worst acceptable response value(s)

When the response is to be maximized, L is the smallest acceptable value, when the response is to be minimized L is the largest acceptable value. When the response is to be on Target, the user gives the smallest and largest acceptable values.

When the response is to be minimized, we must have $T < L$ and when the response is to be maximized we must have $T > L$.

For responses to be on target the user must supply a lower and upper limits such as

$L_l < T < L_u$

L is generated internally when not supplied by the user.

P= The worst response value computed from the starting simplex. P is never closer to the Target than the L(s).

λ is a scaling parameter computed as follows:

$$\lambda = -\ln\left(\frac{\dfrac{100}{(100 - Limit)}}{100 * \dfrac{(L - P)}{(T - P)}}\right)$$

Here Limit = $90 + 80*\text{Log}_{10}(w)$, and w = the weight (importance) assigned to each response by the user. The weight w is a number between 0.1 and 1 (default value).

This definition of λ makes $f(g(y)) = -$ (Limit) when y = L and gives the exponential function $f(g(y))$ the theoretical range: 0 to -100 (this latter limit can only be reached asymptotically when y gets close to Target).

When the user wants the response to be on target, L_u and L_l are respectively used in λ when y $>$T and Y $<$ T.

Overall Desirability

The overall desirability is a weighted average of the individual desirability function. The weights are the **w's,** user entered weights between 0.1 and 1, reflecting the relative importance of the responses.

Overall Distance to Target

An "Overall distance to Target", **D**, used for display is computed as follows:

$$D = \log_{10}\left[\frac{\sum w_i \left(\dfrac{y_i - T}{T - L}\right)^2}{M}\right]$$

Here D = 0 when the y's are all between T and L

M = number of responses

D is computed for display purpose only, and is not used for optimization.

Starting Simplexes

The optimizer starts 8 simplexes from 8 starting runs, selected from 4 corners of the experimental region, the overall center, plus the 3 "best" runs from the Worksheet.

The user can modify these runs or add his own.

Each simplex is generated from the starting run by adding an additional run for each factor with an offset of 20% of its range, the other factors being kept at the same values. A check is made that all runs are within the defined experimental region.

References

A foreword

This is a list covering a small selection of useful references (books and articles) in the fields of design of experiments (DOE) and multivariate data analysis (MVDA). It is emphasized that this is by no means an exhaustive account of the available literature. Rather, this compilation highlights references which may guide the reader for further studies.

References for DOE

Books

1. Box G.E.P., Hunter W.G., Hunter J.S., *Statistics for Experimenters,* John Wiley & Sons, Inc., New York, (1978).

2. Cornell J.A., *Experiments with mixtures,* John Wiley & Sons, Inc., New York, (1981).

3. Bayne C. K., Rubin I.B., *Practical Experimental Designs and Optimization Methods for Chemists,* VCH Publishers, Inc., Deerfield Beach, Florida, (1986).

4. Box G.E.P., Draper N.R., *Empirical Model-Building and Response Surfaces,* John Wiley & Sons, Inc., New York, (1987).

5. Haaland P.D. *Experimental designs in biotechnology*, Marcel Dekker, Inc., New York, Basel (1989).

6. Carlson R., *Design and Optimization in Organic Synthesis,* Elsevier science publishers, Amsterdam (1991).

7. Montgomery, D.C., *Design and Analysis of Experiments*, John Wiley & Sons, New York (1991) ISBN 0-471-52994-X.

8. Morgan E. Chemometrics: Experimental Design, John Wiley & Sons, N.Y., (1991).

9. Nortvedt R et al., *Anvendelse av kjemometri innen forskning og industri,* Tidskriftsforlaget Kjemi AS (1996) ISBN 82-91294-01-1.

Articles

1. Hendrix, C. (1979), *What Every Technologist Should Know About Experimental Design*, Chemtech, 9, 167-174.

2. Hunter, J.S. (1987), *Applying Statistics to Solving Chemical Problems*, Chemtech, 17, 167-169.

3. Grize, Y.L. (1995), *A Review of Robust Process Design Approaches*, Journal of Chemometrics 9, 239-262.

4. Ahlinder, S., et al. (1997), *Smart Testing – Reaping the Benefits of DOE*, Volvo Technology Report No 2 1997, www.volvo.se/rt/trmag/index.html.

5. Nyström, A. and Karlsson, A. (1997) *Enantiomeric Resolution on Chiral-AGP with the aid of Experimental Design. Unusual Effects of Mobile Phase pH and Column Temperature*, Journal of Chromatography A, 763, 105-113.

6. Eriksson, L., Johansson, E., Wikström, C. (1998), *Mixture Design – Design Generation, PLS Analysis and Model Usage*, Chemometrics and Intelligent Laboratory Systems, 43, 1-24.

7. Lundstedt, T., et al. (1998), *Experimental Design and Optimization*, Chemometrics and Intelligent Laboratory Systems, 42, 3-40.

8. Rappaport, K.D., et al. (1998), *Perspectives on Implementing Statistical Modeling and Design in an Industrial/Chemical Environment*, The American Statistician, May 1998, 52, 152-159.

References for MVDA

Books

1. Jollife, I.T. (1986), *Principal component analysis*, Springer-Verlag, New York (ISBN 0-387-96269-7).

2. Martens, H. and Naes, T. (1989), *Multivariate calibration*, John Wiley, New York.

3. Jackson, J.E. (1991), *A user's guide to principal components*, John Wiley, New York. (ISBN 0-471-62267-2).

4. Anthology (1996), *Anvendelse av Kjemometri innen forskning og industri*, Tidsskriftforlaget Kjemi AS, Bergen, Norway (ISBN 82-91294-01-1).

5. Höskuldsson, A. (1996), *Prediction Methods in Science and Technology,* Thor Publishing, Copenhagen, Denmark (ISBN 87-985941-0-9).

6. Massart, D.L., *et al.*, (1997 & 1998) *Handbook of Chemometrics and Qualimetrics*. Part A and B. Elsevier, Amsterdam.

Articles, general

1. Wold, S., Esbensen, K., Geladi, P. (1987), *Principal Component Analysis*, Chemometrics and Intelligent Laboratory Systems, 2, 37-52.

2. Höskuldsson, A. (1988), *PLS Regression Methods*, Journal of Chemometrics, 2, 211-228.

3. Ståhle, L., and Wold, S. (1988), *Multivariate Data Analysis and Experimental Design in Biomedical Research*, In: Ellis, G.P., and West, G.B. (Eds) *Progress in Medical Chemistry*, Elsevier Science Publishers, 291-338.

4. Wold, S., Albano, C., and Dunn W.J., et al. (1989), Multivariate Data Analysis: Converting Chemical Data tables to plots, In: *Computer Applications in Chemical Research and Education,* Heidelberg, Dr. Alfred Hütig Verlag.

5. Stone M., Brooks R.J. (1990), *Continuum regression: Cross-validated sequentially constructed prediction embracing ordinary least squares, partial least squares and principal components regression*, Journal of the Royal Statistical Society, Ser. B, 52, 237-269.

6. Frank, I.E., and Friedman, J.H. (1993), *A Statistical View of Some Chemometrics Regression Tools*, Technometrics, 35, 109-148.

7. Wold, S. (1994*), Exponentially Weighted Moving Principal Components Analysis and Projections to Latent Structures*, Chemometrics and Intelligent Laboratory Systems, 23, 149-161.

8. Wold, S., Eriksson, L., and Sjöström, M. (1999), PLS in Chemistry, In: The Encyclopedia of Computational Chemistry, Schleyer, P. v. R.; Allinger, N. L.; Clark, T.; Gasteiger, J.; Kollman, P. A.; Schaefer III, H. F.; Schreiner, P. R., Eds., John Wiley & Sons, Chichester, pp. 2006-2020.

Articles, process

1. Kresta, J.V., MacGregor J.F., and Marlin T.E. (1991*), Multivariate Statistical Monitoring of Process Operating Performance*, The Canadian Journal of Chemical Engineering, 69, 35-47.

2. Kourti, T., and MacGregor, J.F. (1995), *Process Analysis, Monitoring and Diagnosis, Using Multivariate Projection Methods*, Chemometrics and Intelligent Laboratory Systems, 28, 3-21.

3. MacGregor, J.F. (1996), *Using, On-line Process Data to Improve Quality*, ASQC Statistics Division Newsletter, vol. 16. NO. 2. Page 6-13.

4. Nijhuis, A., de Jong, S., Vandeginste, B.G.M. (1997), *Multivariate Statistical Process Control in Chromatography,* Chemometrics and Intelligent Laboratory Systems, 38, 51-61.

5. Rännar, S., McGregor, J.F., and Wold, S. (1998), *Adaptive Batch Monitoring Using Hierarchical PCA,* Chemometrics and Intelligent Laboratory Systems, 41, 73-81.

6. Wikström, C., et al. (1998), *Multivariate Process and Quality Monitoring Applied to an Electrolysis Process – Part I. Process Supervision with Multivariate Control Charts*, Chemometrics and Intelligent Laboratory Systems, 42, 221-231.

7. Wikström, C., et al. (1998), *Multivariate Process and Quality Monitoring Applied to an Electrolysis Process – Part II. Multivariate Time-series Analysis of Lagged Latent Variables*, Chemometrics and Intelligent Laboratory Systems, 42, 233-240.

8. Wold, S., et al. (1998), *Modelling and Diagnostics of Batch Processes and Analogous Kinetic Experiments,* Chemometrics and Intelligent Laboratory Systems, 44, 331-340.

Articles, multivariate calibration

1. Brown, P.J. (1982), *Multivariate Calibration,* Journal of the Royal Statistical Society, B44, 287-321.

2. Beebe, K.R. and Kowalski, B.R. (1987), *An Introduction to Multivariate Calibration and Analysis*, Analytical Chemistry, 57, 1007-1017.

3. Trygg, J., and Wold, S. (1998), *PLS Regression on Wavelet Compressed NIR Spectra,* Chemometrics and Intelligent Laboratory Systems, 42, 209-220.

4. Swierenga, H., et al. (1998), *Improvement of PLS Model Transferability by Robust Wavelength Selection,* Chemometrics and Intelligent Laboratory Systems, 41, 237-248.

5. Wold, S., Antti, H., et al. (1998), *Orthogonal Signal Correction of Near-Infrared Spectra,* Chemometrics and Intelligent Laboratory Systems, 43, 123-134.

6. Bro, R. (1996), Håndbog i Multivariabel Kalibrering, KVL, Copenhagen, Denmark.

Articles, multivariate characterization

1. Carlson, R., Lundstedt, T., and Albano, C. (1985), *Screening of Suitable Solvents in Organic Synthesis – Strategies for Solvent Selection*, Acta Chemica Scandinavica, B39, 79-91.

2. Wallbäcks, L., Edlund, U. and Nordén, B. (1991), *Multivariate Characterization of Pulp Using Solid-State ^{13}C NMR, FTIR and NIR,* Tappi Journal, 74, 201-206.

3. Cocchi, M., et.al. (1992), *Theoretical versus Empirical Molecular Descriptors in Monosubstituted Benzenes – A Chemometric Study,* Chemometrics and Intelligent Laboratory Systems, 12, 209-224.

4. Eriksson, L., Verhaar, H.J.M., and Hermens, J.L.M. (1994), *Multivariate Characterization and Modelling of the Chemical Reactivity of Epoxides*, Environmental Toxicology and Chemistry, 13, 683-691.

5. Lindgren, Å., and Sjöström, M. (1994), *Multivariate Physico-Chemical Characterization of Some Technical Non-Ionic Surfactants,* Chemometrics and Intelligent Laboratory Systems, 23, 179-189.

6. Andersson, P., Haglund, P., and Tysklind, M. (1997), *Ultraviolet Absorption Spectra of all 209 Polychlorinated Biphenyls Evaluated by Principal Component Analysis,* Fresenius Journal of Analytical Chemistry, 357, 1088-1092.

Articles, QSAR

1. Eriksson, L., Hermens, J.L.M., et al. (1995), *Multivariate Analysis of Aquatic Toxicity Data with PLS,* Aquatic Sciences, 57, 217-241.

2. Eriksson, L., and Johansson, E. (1996), *Multivariate Design and Modeling in QSAR*, Chemometrics and Intelligent Laboratory Systems, 34, 1-19.

3. Verhaar, H.J.M., Hermens, J.L.M., et al. (1996), *Classifying Environmental Pollutants. Separation of Class1 and Class2 Type Compounds Based on Chemical Descriptors*, Journal of Chemometrics, 10, 149-162.

4. Goodford, P. (1996), *Multivariate Characterization of Molecules for QSAR Analysis,* Journal of Chemometrics, 10, 107-117.

5. Lindgren, Å., et al. (1996), *Quantitative Structure-Effect Relationships for Some Technical Non-ionic Surfactants,* Journal of the American Oil Companies Society, 73, 863-875.

6. Sandberg, M., et al. (1998), New Chemical Descriptors Relevant for the Design of Biologically Active Peptides. A multivariate Characterization of 87 Amino Acids, Journal of Medicinal Chemistry, 41, 2481-2491.

Index

2

2^2 full factorial design 60, 153, 216
2^3 full factorial design 62, 78, 119–20, 153, 155, 202
2^4 full factorial design 63
2^5 full factorial design 63

A

Ability to handle a failing corner 74
Action limits in the N-plot 100
Adding complementary runs 132, 150
Advanced mixture designs 236
Algebraic approach to D-optimality 216
Analysis of factorial designs 85, 93, 103, 110
Arranging factors in inner and outer arrays 248
Automatic search for an optimal point 142, 182
Averaging procedure in DOE 19
Axial experiments 14, 171–72, 174

B

Bad replicates 108–9, 112
Balanced distribution of experiments 120
Balancing 43, 50, 222
Benefits of response transformation 105
Box-Behnken designs 191–93, 196
Box-Whisker plot 98, 103–5

C

Candidate set 216–17, 220, 225, 230–31, 307–8, 315–16, 318

Cause of poor model
    Skew response distribution 103
    Curvature 107
    Bad replicates 108
    Deviating experiments 109
CCC design in three factors 171–73
CCC design in two factors 170–71
CCF design in three factors 172–73
Central composite circumscribed 165, 170
Central composite face-centered 165, 169, 172
Classical analysis approach 247, 249–50, 253, 255, 257, 263
Coding of qualitative variables 242
Coincidental noise effect 20
Combining process and mixture factors 223
Complementary half-fraction 153–54
Complementing for unconfounding 132
Composite design family 14, 165
Computation of effects using least squares fit 74
Computation of main effects 68
Computational matrix 70, 120, 151–52, 216
Conceptual basis of semi-empirical modelling 26
Condition number 93–94, 98, 101, 110, 112, 125, 135, 170, 175, 205, 213, 220–22, 226–27, 270, 279, 291, 293–94, 297–99, 317
Confidence interval 80–81, 101, 209, 265, 267–71, 295, 299
Confidence limits 268
Confounding of effects 121, 123–24, 149, 158
Confounding pattern 121–22, 151–52, 154–59, 163
Connecting factors and responses 23
Consequence of variability 21–22
Controllable factors 35, 165
COST approach 2–3
Creation of experimental worksheet 40
Cubic model 39, 182, 195, 307, 309
Curvature 55, 88, 105, 107, 112, 132, 173, 210, 302, 318
Curved plane 57

D

Defining relation 151, 155–59, 163
Definition of a two-factor interaction 53
Deleted studentized residuals 109–10, 299
Design factors 247–49, 253, 263
Deviating experiments 99, 102, 106, 109, 112, 253, 263
D-optimal approach 194, 213, 216–18, 220, 227
D-optimal design 46–48, 50, 94, 132, 151, 159, 161–63, 170, 191, 194–96, 213–14, 216, 220, 22-223, 225, 227, 236, 238, 245, 307, 313–17, 318

E

Easy-to-control factors 247
Empirical models 24, 35, 51
Estimable terms 119
Estimation of the experimental noise 74
Evaluation of raw data
  Condition number 93
  Descriptive statistics of response 97
  Histogram of response 96
  Replicate plot 85
  Scatter plot 94
Expensive and inexpensive factors 262
Experimental objective
  Optimization 165, 191
  Robustness testing 197
  Screening 113, 151
Experimental stage: optimization 33
Experimental stage: screening 31
Expressing regression coefficients 79
Extended coefficients display 243
Extrapolation 140–41, 143–46, 149, 182, 184–85, 189, 310
Extreme vertex 237

F

Factor transformation 48–49
Familiarization stage 30, 35, 37
F-distribution 265, 271–72
Features of the D-optimal approach 220
Finding the optimal region 29–30, 32, 42, 115, 139–40
Fitting a hyperplane 52
Fitting of a special regression model 162

Focusing on effects 19
Fold-over design 133, 220
Fractional factorial design family 14, 159
Fractional Factorial Designs 14, 39, 59, 66, 113, 118–19, 122–23, 132, 149–50, 151, 155, 159–61, 163, 170, 174, 188, 201–3, 244, 293
Framework for experimental planning 14
Full factorial designs 5, 44, 59, 63, 65–66, 75, 85, 92, 112, 120, 123, 132, 149, 159, 174, 191, 193, 196, 245

G

G-efficiency 213, 220–22, 226–27
General Example 1: Screening 9
General Example 2: Optimization 10
General Example 3: Robustness testing 11
Generation of alternative D-optimal designs 225
Generation of design 39, 42, 118, 169, 201
Generation of mixture design 231
Generators 151, 153–57, 159, 163, 293
Geometric approach to D-optimality 217
Geometric representation of fractional factorial designs 119
Geometry of interaction models 53
Geometry of linear models 51
Geometry of quadratic models 54
Global model 26, 28
Goodness of fit 87, 92, 276–77, 295–97
Goodness of prediction 87, 92, 104, 277
Gradient techniques 33, 35, 140–41, 149–50
Graphical evaluation of optimization results 186
Graphical interpretation of confoundings 122

H

Hard model 25
Hard-to-control factors 247
High condition number 110, 112, 279, 291
Histogram of response 93, 96, 98, 102, 106–7, 270
How to compute a determinant 218

I

Inclusion of existing experiments 162, 195
Influence of noise 17, 265
Informationally optimal experiments 19
Inner and outer arrays 247–49, 263

Q

Quadratic model 33, 39, 42, 54–57, 95, 147, 162, 167–69, 171, 173, 176, 182, 189, 195, 234, 257, 259, 308
Quadratic polynomial models 23, 33, 38, 57, 166
Qualitative factor 36, 41, 43–44, 53, 194, 201, 222, 239, 242–45, 257, 298
Qualitative factors at many levels 44–45, 50, 162, 239
Qualitative factors at two levels 43–44, 195, 239
Quantitative factor 21, 36–37, 41, 44, 53, 239, 244
Quarter-fraction 133, 155, 202

R

R^2/Q^2 pointing to a poor model 87
Randomized run order 41, 118
Raw residuals 109, 298
Reacting to noise 18
Real effects 20–22
Reference mixture 223, 233, 237, 308–12
Regression analysis
    Normal probability plot of residuals 99
    The R^2/Q^2 diagnostic tool 86
    Model significance test 274
Regular and extended lists of coefficients 243
Regular coefficients display 243
Regular experimental region 46
Regularly arranged corner (factorial) experiments 170
Relationships between factors and responses 4, 25, 95, 166
Repeated testing of the standard condition 12
Replicated center-points 30–31, 88, 117–18, 134, 153, 160–61, 170–71, 174, 188, 191, 203, 210
Resolution of fractional factorial designs 122, 159
Resolution III 123, 132–33, 158, 161, 201–3
Resolution IV 123, 133, 149, 158–59, 163, 203
Resolution V 123, 133–34, 158, 171
Resources for a fixed number of runs 162
Response surface methodology 170
Response surface modelling 33, 166, 170, 181, 191
Robust design 247–48, 261, 263
    Connotation and objective 247

Robustness testing
    Introduction 34
    Inside specification/Non-significant model 207
    Inside specification/Significant model 206
    Outside specification/Non-significant model 210
    Outside specification/Significant model 208
Rotatability 193
Run order 29, 41, 233

S

Scaled and centered regression coefficients 80, 235
Screening large numbers of factors 149
Selection of experimental objective 42, 115, 167, 199, 230
Selection of model 38
Selection of regression model 39, 42, 117, 168, 201
Semi-empirical modelling 25–26, 28
Semi-empirical models 24, 35, 51
Set point 197, 199
Several correlated responses 103, 110, 112, 279, 291
Simplex methodology 142, 183
Simplifying the response function 105
Simultaneous start of eight simplexes 143, 183
Six stages in DOE 42
Size and shape of mixture region 232
Skew response distribution 103, 112
Soft model 25
Specification of factors 35–36, 42, 115, 167, 199
Specification of responses 37, 42, 117, 168, 200
Specifying constraints on factors 46, 47
Stabilizing the variance of the residuals 105
Stages in the experimental process 29–30, 35
Standard operating condition 8, 12
Standard reference experiment 8
Standardized residuals 109, 299
Steepest ascent 32–33, 140
Steepest descent 32, 140
Sum of squares 78, 273–75, 290, 296, 300–301
Surfaces mimicked by quadratic models 169
Symmetrically arrayed star points 171, 188
Systematic and unsystematic variability 17

T

Taguchi approach to robust design 247, 263
Taylor series expansions 25

U

V

W