◆ UMETRICS

Contents

3 Factorial designs   53

4 Analysis of factorial designs 75

5 Experimental objective: Screening 101

6 Experimental objective: Optimization   145

7 Experimental objective: Robustness Testing

169

8 Additional screening designs for regular regions

183

9 Additional optimization designs for regular regions 201

10 D-optimal design 217

14 Blocking the experimental design 279

15 The Taguchi approach to robust design 293

16 Models of variability: Distributions 309

Index   453

1 Introduction

1.1 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 text-book.

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 chapter 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. 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.

Additionally, 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. 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.

Finally, 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.

1.2 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

1.3 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*.
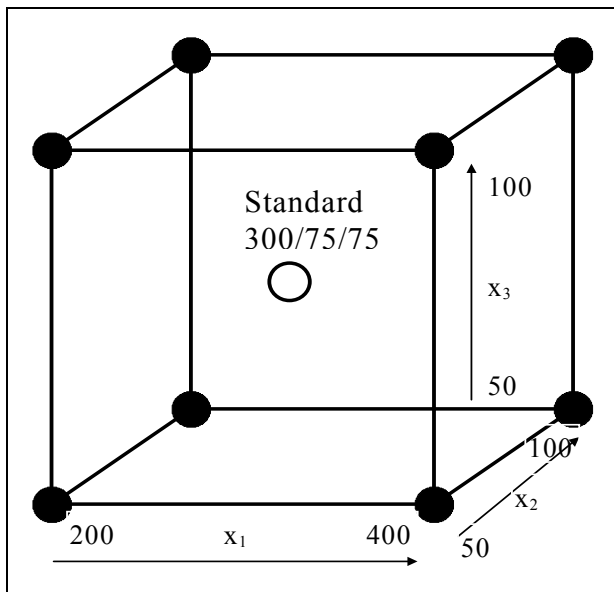
Figure 1.1: A symmetrical distribution of experimental points around a center-point experiment.

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.

1.4 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.
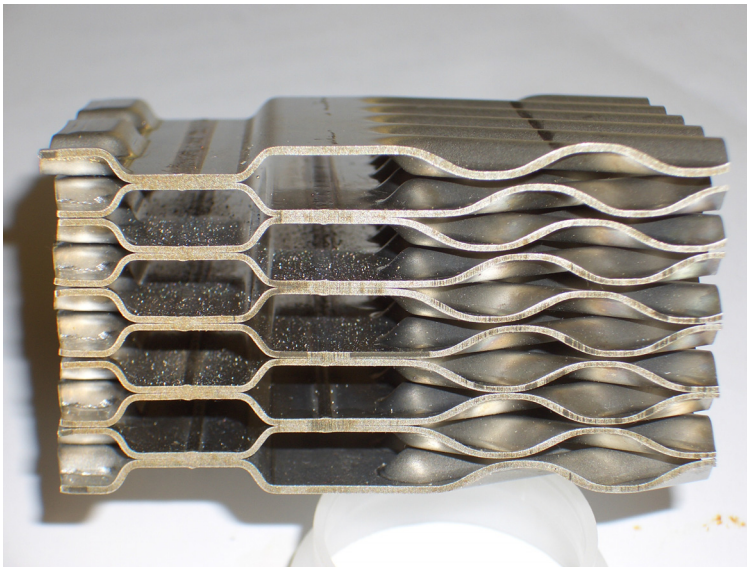
Figure 1.2: A cross-section of a plate heat-exchanger developed and manufactured by Alfa Laval Thermal.

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.

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.

1.5 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 6.

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.

1.6 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).

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, the capacity factors measure the retention of compounds, and the 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 7).

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.

1.7 A small example: The CakeMix data set

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.
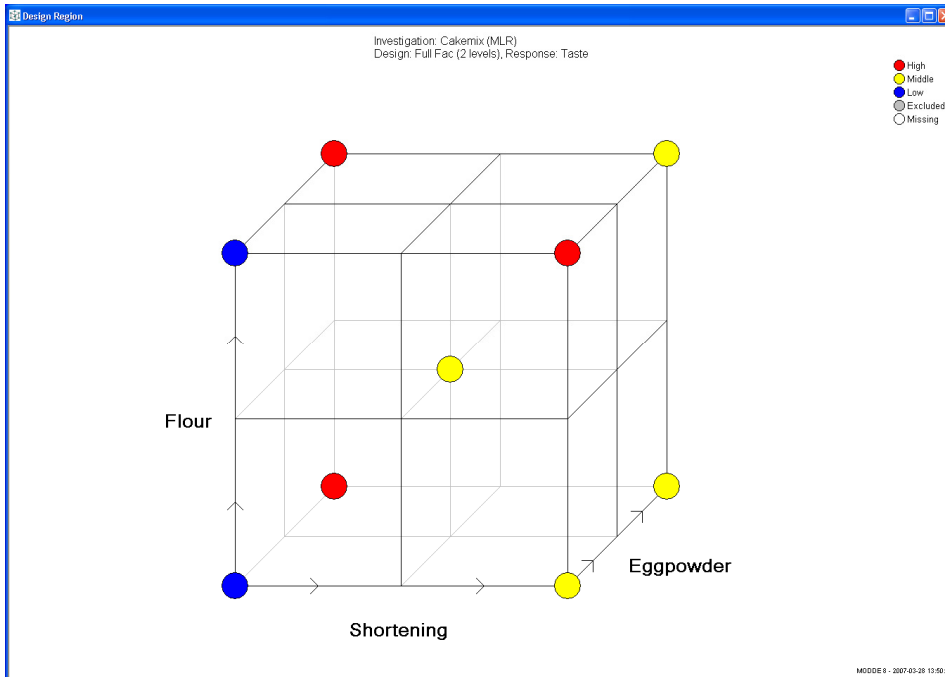
Figure 1.10: *A geometrical representation of the CakeMix experimental protocol.*

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 modeling, and model adequacy are further discussed in Sections 1.15 – 1.17.

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.
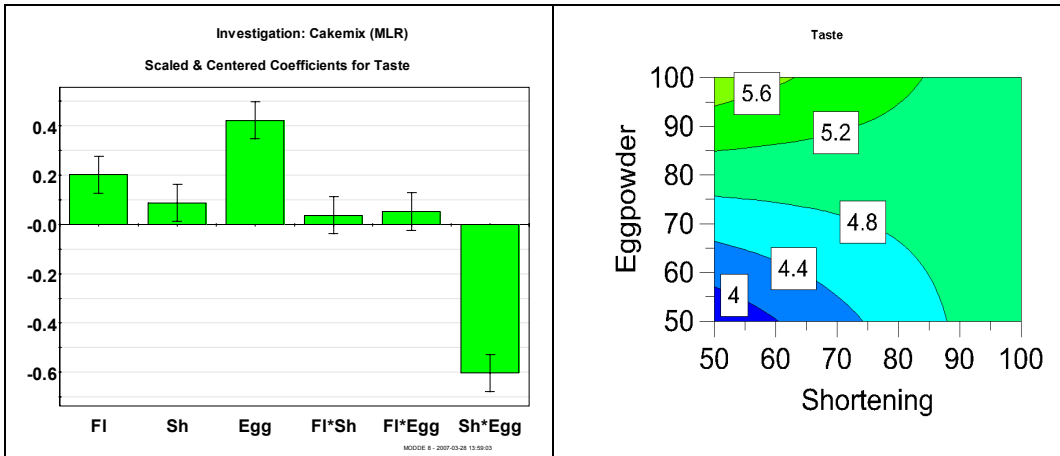
Figure 1.11: *(left) Regression coefficient plot of CakeMix regression model.*
Figure 1.12: *(right) Response contour plot of taste as a function of Eggpowder and Shortening.*

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 modeled 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.

1.8 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.

1.9 Variability

We will now discuss the concept of variability. Consider Figure 1.13 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
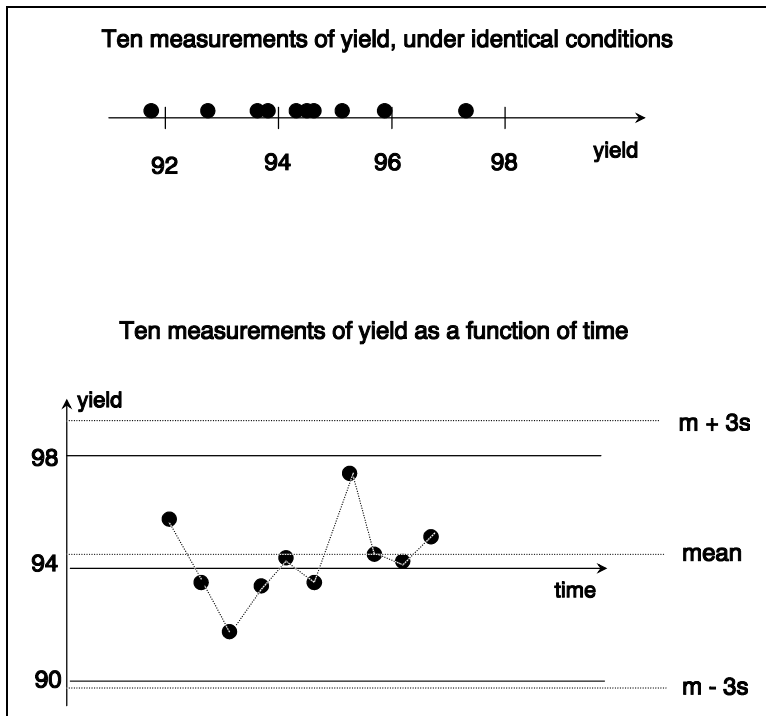
Figure 1.13: The yield of a product measured ten times under identical experimental conditions.

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 1.13, 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.

1.10 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 1.14. 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 ten 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.
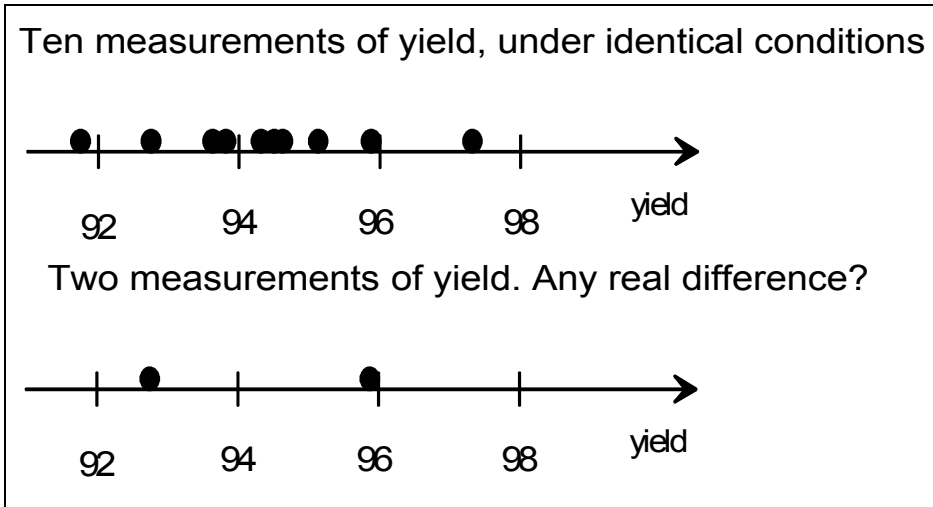
Figure 1.14: *An illustration of the effect on yield obtained by changing the settings of one critical factor.*

1.11 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 1.15. 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.
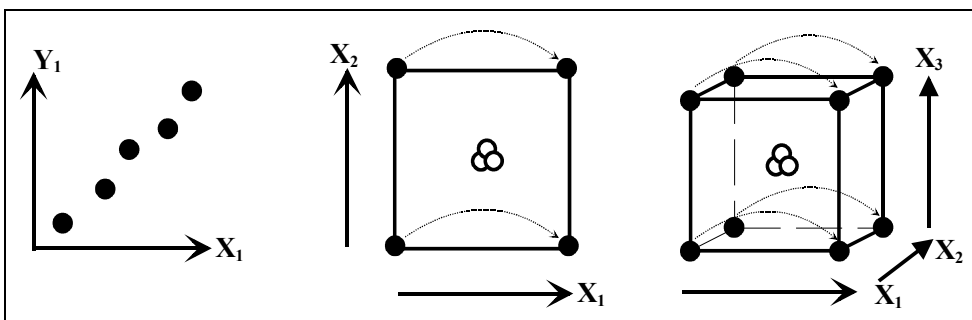
Figure 1.15: *The averaging procedure in DOE leading to more precise effect estimates.*

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 the factors, and not on some coincidental noise effect. In addition, DOE always needs fewer experiments than COST.

1.12 Graphical display of real effects and noise

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 1.16 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.*
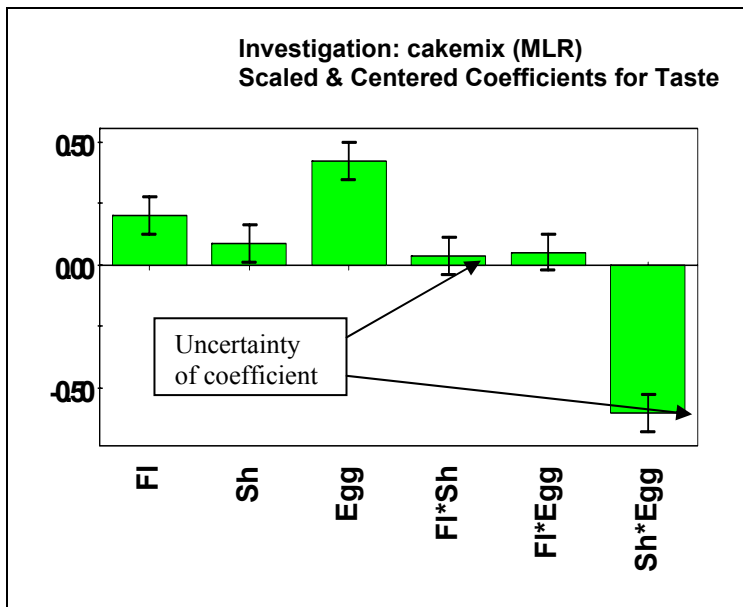
Figure 1.16: *Regression coefficients of the CakeMix interaction model.*

We now turn to the interpretation of the model in Figure 1.16. 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 Eggpowder. 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.

1.13 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 1.17. 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 1.17), 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 1.17. 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?
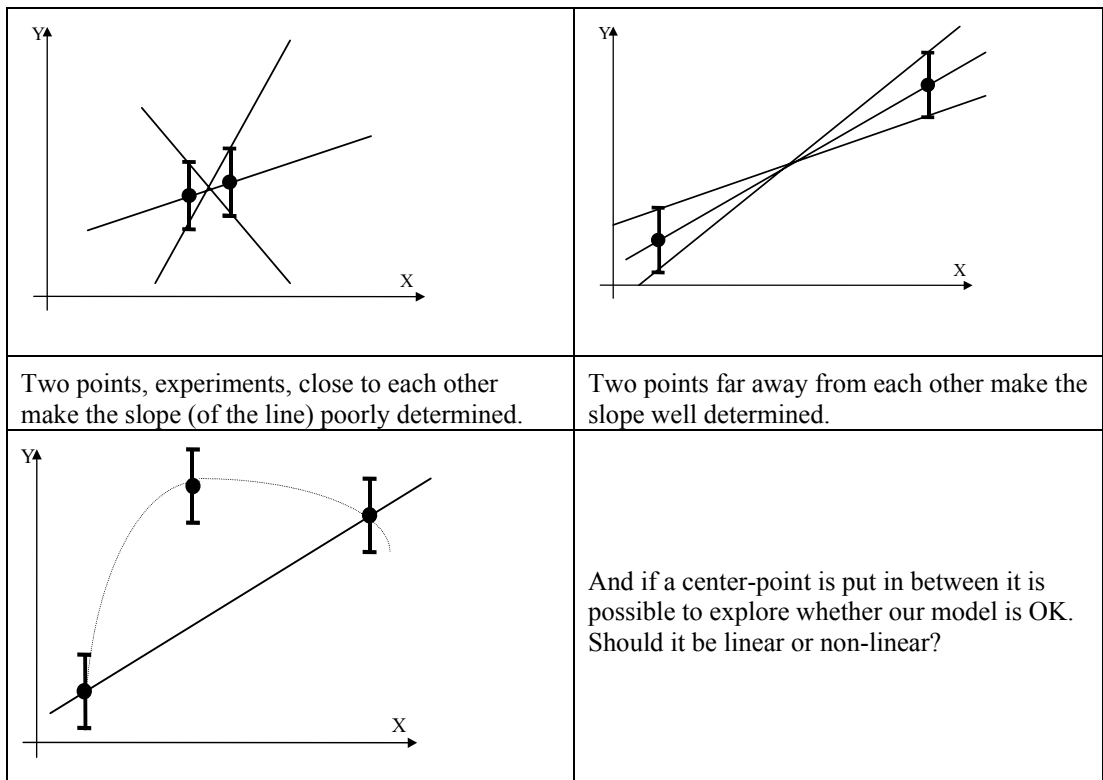
Figure 1.17: *A consequence of variability is that it matters where the experiments are performed.*

1.14 Connecting factors and responses

In DOE, there are two fundamental types of variables, *factors* and *responses* (see Figure 1.18). 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 a cake mix recipe, gas flow at nozzle of a welding equipment, exhaust gas re-circulation in a 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.
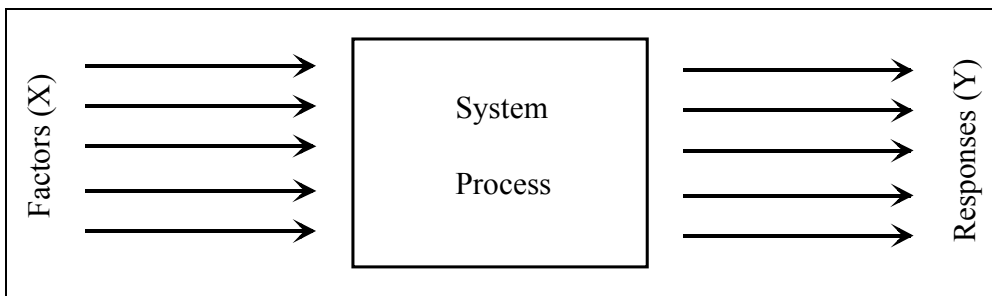
Figure 1.18: The measured responses describe the properties of the investigated system. By changing the most influential factors the features of the system might be altered according to a desired response profile.

1.15 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 1.19. 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.
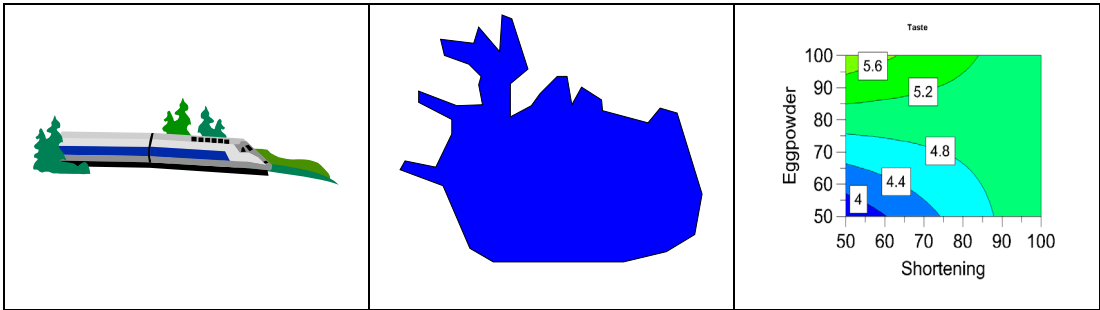
Figure 1.19: *Three examples of models, a toy train, a map of Iceland, and a response contour plot of taste.*

1.16 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 1.20). 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 enter this kind of prior knowledge into an empirical investigation, we are conducting *partial* empirical modeling, and consequently the prefix semi- is often added. Because of this, it is often stated that DOE involves semi-empirical modeling 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 1.20: *An overview of empirical, semi-empirical and fundamental mathematical models.*

1.17 Semi-empirical modeling – Taylor series expansions

Taylor series expansions are often used in semi-empirical modeling. Consider Figure 1.21 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: $y = P(x) = b_0 + b_1x + b_2x^2 + ... + b_px^p + e$, where e represents a

residual term. In this polynomial function, the degree, p, gives the complexity of the equation.
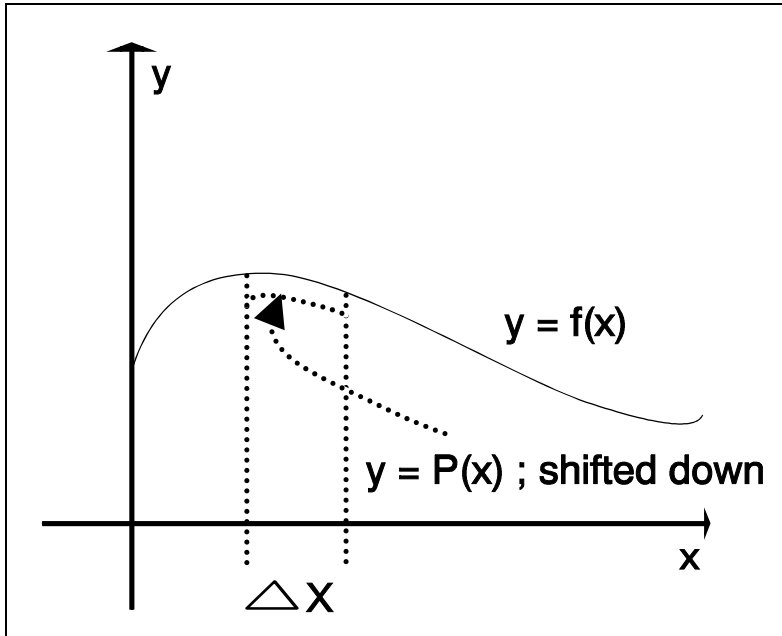
Figure 1.21: *The "true" relationship, y = f(x), between a response, y, and a factor, x, approximated by an alternative function, y = P(x), in a limited factor interval Δx.*

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 1.21. 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 modeling 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.

1.18 Conceptual basis of semi-empirical modeling

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

---

1.22). 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.
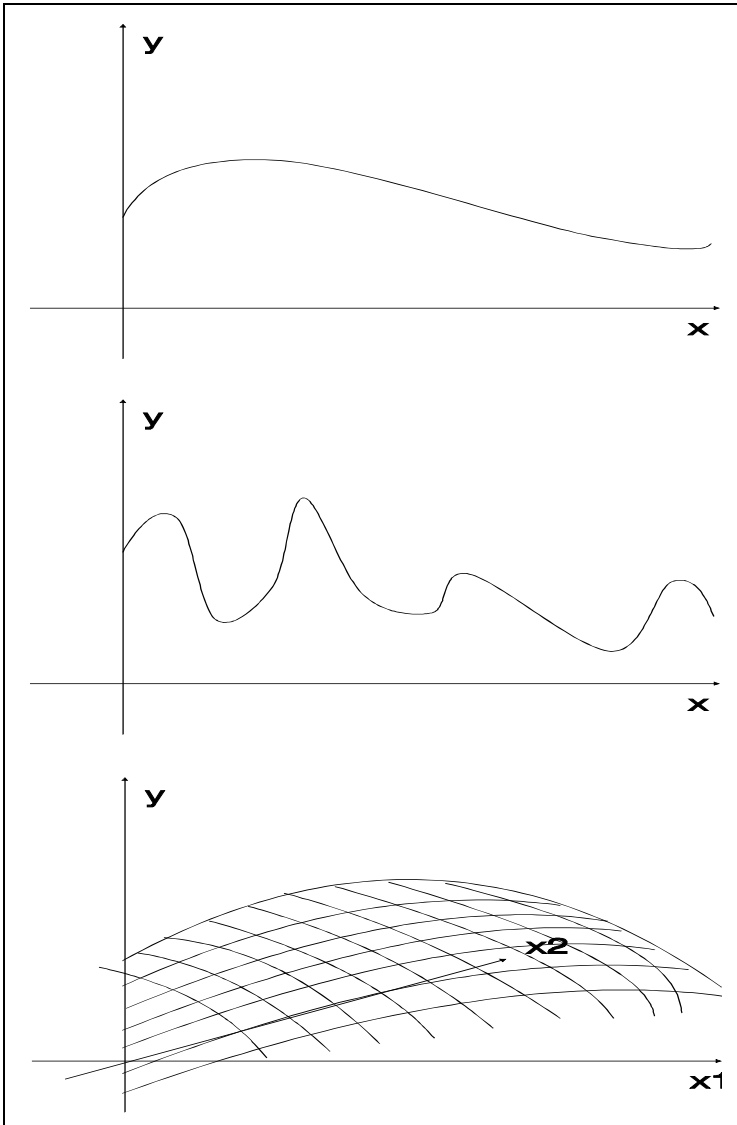
Figure 1.22: *Our long experience (technology, chemistry, biology, medicine,...) shows that nature is fairly smooth and not rugged. With several factors a smooth response surface is usually applicable. Such a smooth surface can be well approximated by a quadratic polynomial model.*

1.19 Configuration of experiments depends on the objective

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 1.23 provides a summary of the designs discussed.

The first row of Figure 1.23 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 1.23 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.

2 factors	3 factors	> 3
		Hyper cube
		Balanced fraction of hyper cube
		Hyper cube + axial points
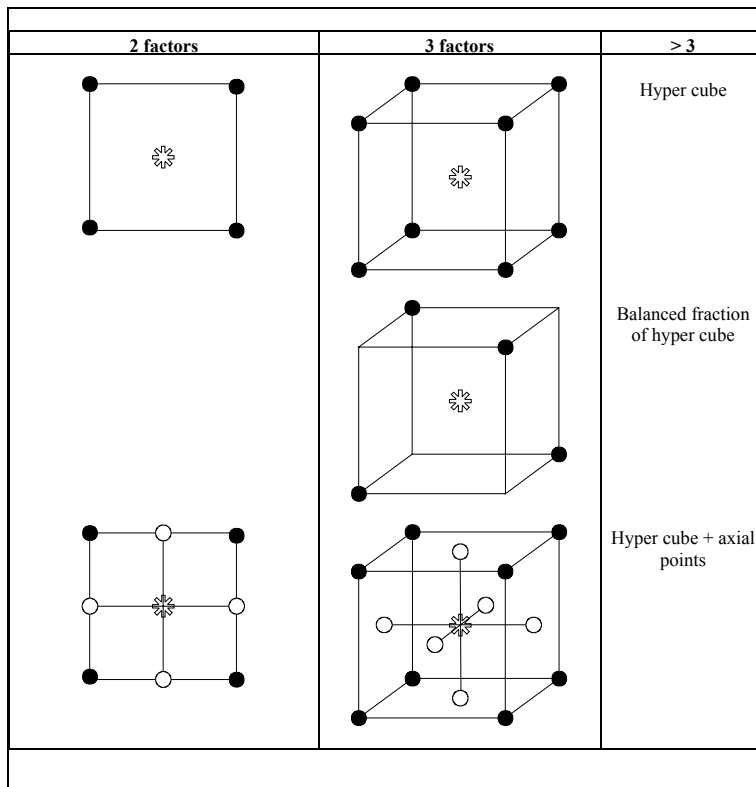
Figure 1.23: Examples of full factorial, fractional factorial, and composite designs used in DOE.

1.20 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.

1.21 Questions for Chapter 1

- Which are the three main experimental objectives?
- What are the main questions addressed using these objectives?
- 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?
- 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?

1.22 Summary and discussion

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.

It was shown that 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.

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 modeling, 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.