EXPERIMENT PLANNING AND THE USE OF FACTORIAL DESIGNS

By G. R. E. LIONNET and S. M. BAKER
Huletts Sugar Limited, Mount Edgecombe

Abstract

The importance of overall planning for efficient experimentation is discussed and stressed. One of the most important steps of this planning, the selection of an experimental design, is studied with particular reference to the use of factorial designs. The advantages of this type of design as well as a recent application are described.

Introduction

This paper deals with four aspects of experimentation.

Firstly, it provides a set of procedures considered essential for efficient experimental planning.

Secondly, it highlights the advantages in using properly planned, statistically designed experiments instead of completing a series of experimental trials and then posing the questions "How do we analyse the data?" "What conclusions can be made from the results?"

Thirdly, it discusses in some detail, an experimental design.

Finally, it gives an example of a problem where a factorial design has been employed as an initial step in the problem solving process.

Procedures for efficient experimentation

As efficient experimentation requires that the maximum useful information be obtained from the minimum outlay of money, time, effort, etc., some basic procedures have to be followed:

1. Define the problem.
2. Obtain background information.
3. Design the experimental programme. In this paper, attention is focused on the design of the experimental programme, as follows:—
   (*a*) Select the independent variables or factors to be studied and establish the values at which they are to be tested. These values are usually called "levels" of the factors.
   (*b*) Choose the dependent variables which are to be measured; the values thus obtained are the results of the experiments. These dependent variables are usually called "response variables".
   (*c*) Consider the probable precision and accuracy of the measurements to be taken and of the final results.
   (*d*) Consider the limitations, e.g. cost, labour, time, materials, etc.
   (*e*) Consider possible interrelationships of the factors, i.e. whether the effect of one factor is dependent on the level of another.
   (*f*) Consider the effects of factors not under study. For example a seasonal change could affect the results if the experiment covers a long time period, in which case the effect of time must be taken into consideration.
   (*g*) Select an experimental design and draw up a detailed chronological plan of action.
4. Carry out the experimental work.
5. Interpret the experimental results.

Achieving efficient experimental work without applying at least part of these procedures would be difficult, if not impossible. It is probable that they are used, at least intuitively, by all experimenters.

One of the most important steps in the above procedure is the selection of the experimental design.

Factorially designed experiments

Factorially designed experiments offer some definite advantages, the most valuable being the determination of interactions between factors. Two factors are said to interact if the effect of one depends on the level of the other. These designs also indicate the relative importance of each factor — thus they may be used for "screening", that is, selecting from a large set of factors those that are important. Statistical designs yield unambiguous results and an estimate of the experimental error at a minimum cost. Finally, it is usually possible to proceed sequentially, which allows constant adaptation of the experimentation in the light of results obtained.

A common approach when investigating the effect of several factors on a response is the "one variable at a time method". In this kind of experiment the effect of varying one factor is measured while holding the others constant. This is repeated so that at each stage a different factor is varied while the others are held constant. The results of such an experiment are incomplete because it is impossible to determine whether the effect of one factor is dependent on the level of another, i.e. there may be an interaction between the two factors, which cannot be detected with this "one at a time" approach. The concept of interaction is presented graphically in Fig. 1.

In a factorial experiment several factors are controlled, at each of 2 or more levels, and their effects on the response variable investigated. The response is measured at each one of all the possible combinations that can be formed for the different levels of the factors. Thus if 4 factors are chosen, each at 2 levels, the experiment will consist of $2 \times 2 \times 2 \times 2 = 16$ trials, i.e. a 2^4 factorial experiment, where 2 represents the number of levels and 4 the number of factors. This design allows not only the effect of each factor on the response to be calculated but also the effect of any interaction of factors. The way in which these are calculated is described in detail in the literature[1].

The magnitude of these effects should be compared to the experimental error inherent in the system, so that it may be judged whether they genuinely are a result of changes in the level of the factors or whether they are merely the effect of experimental error. Thus an estimate of the experimental error is required for the proper analysis of a factorial design. If there is no idea of the size of the experimental error from previous experimentation it may be estimated by using either replication (certain or all trials are repeated and the variance of the repeated observations determined) or otherwise by combining the higher order interaction effects as an estimate of the experimental error. This last possibility may only be used if it can be assumed that these higher order interactions are negligible, which is often the case.

Fractional factorial experiments

A complete factorial experiment in which all possible combinations of all the levels of the different factors are investi-

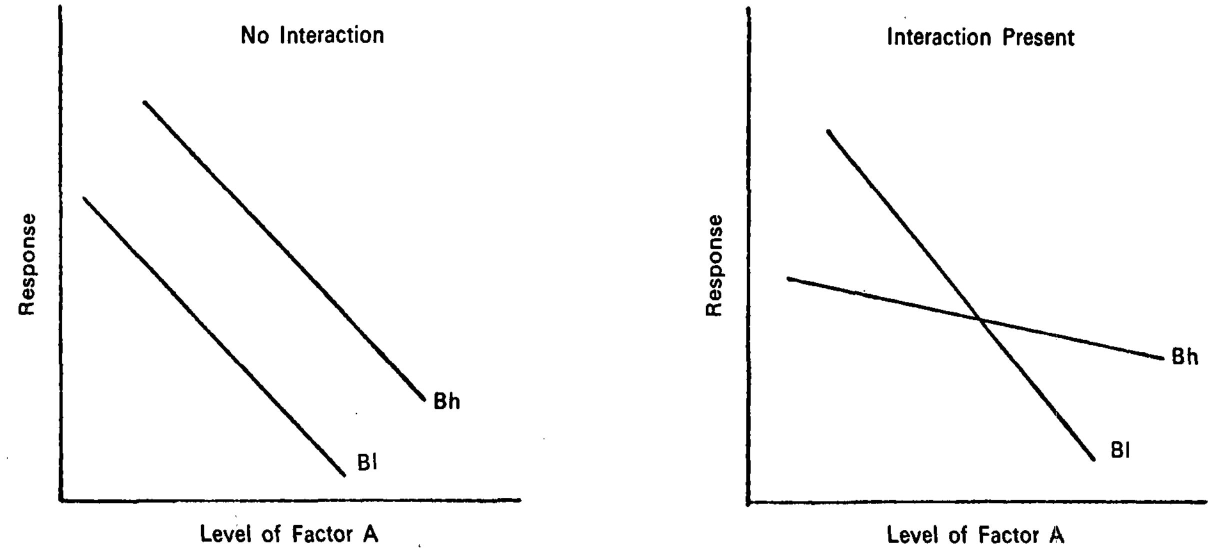

FIGURE 1 Effect of Different levels of A on the Response for Two Different levels of B.

gated is often out of the question because of the large number of runs required. For instance, a factorial experiment having 7 or 10 variables, each at two levels, will require 128 or 1 024 runs, respectively. Fortunately, it is possible to investigate the main effects of the factors and their more important interactions in a fraction of the total number of runs required for the complete factorial. The appropriate experimental designs are termed fractional factorials and are described in detail elsewhere[1].

Confounding and randomisation in factorial designs

Factors which might affect the outcome of an experiment but which the experimenter is not controlling should not be varied together with the factors under study. If variation in such background factors cannot be avoided, the technique known as "confounding" or "blocking" may be employed to deal with this situation[1, 6]. Another method of ensuring that such background factors do not annul the results of the experiment is by randomisation[1, 6]. In this approach the sequence of tests to be run is done in a random order. The effect of the uncontrolled factors is thus lumped together into the experimental error as unaccounted variability. Although randomisation is the simpler of the two procedures to apply, its disadvantage over confounding is that the effect of background variables is not removed from the experimental error.

Example of a factorial design

Although factorial designs are a frequent means of experimentation in industrial research[3, 4, 5, 7, 9, 10] they appear to be little utilised in the sugar industry[2]. The authors have applied this technique in a number of projects, the most ambitious being the optimisation of C-massecuite exhaustion involving the use of a pilot-plant crystalliser.

Of the factors possibly affecting C-massecuite exhaustion, the following were considered the most important and worthy of further study:

Massecuite brix at pan drop.
Massecuite apparent purity.
Stirring rate in crystallisers.
Cooling rate in crystallisers.
Retention time in crystallisers.
Type and quantity of impurities in the massecuite.

While massecuite brix and purity may be readily controlled on a factory scale stirring, cooling and retention in crystallisers are only easily regulated on the experimental scale. Unfortunately nothing can be done to control the last factor, which is thus a limitation to the experimental programme.

It was assumed that over a period of a month effects due to changes in massecuite impurities would not seriously affect information obtained about the factors of interest. Thus the aim of the first experiment should be to investigate the influences of the above 5 factors on C-massecuite exhaustion and to establish the validity of the assumption concerning impurities.

Design

Since the investigation involves a number of factors which are likely to interact, a factorial design, consisting of all possible combinations of the factors each at 2 selected levels, i.e. 5 factors at 2 levels each which yields 2^5 or 32 runs, would be a logical choice.

However, as it was impossible to complete 32 runs in a month (to which should be added a number of replicate runs for the measurement of experimental error) cooling rate was eliminated from this first programme since it may be linked to retention (i.e. a long retention implying a slow cooling rate and vice-versa). Two standard cooling rates were thus established and used.

This resulted in 4 factors at 2 levels or a 2^4 factorial which requires 16 runs. At 5 runs per week, that is 4 of the set runs plus 1 replicate run, 20 runs could be done in a month.

Two important points in level selection are firstly that the levels of the factors should be within the bounds of what could be reasonably achieved under factory conditions and, secondly, to improve experimental precision the levels should be as far apart as practically permissible. Bearing this in mind, the levels shown in Table 1 were chosen.

TABLE 1
Levels of the factors

Factor	Abbreviation	Low level (*l*)	Normal operating level	High level (*h*)
Massecuite brix at pan drop	B	94	98	97
Massecuite apparent purity	P	48 to 52*	52	56 to 60*
Retention (hrs)	R	6	40	10
Stirring rate (rpm)	S	3	—	6

* As it was impossible to obtain rigid values for purity two ranges were set.

Execution of test programme

Using the abbreviations shown, e.g. B*l*, P*l*, S*l*, R*h*, which mean low brix, low purity, low stirring and high retention, the 16 runs were established. In addition to these, one of the combinations B*l*, P*l*, S*h*, R*l* was chosen as the replicate run.

To prevent the association of possible time effects with one particular factor, the runs were carried out in random order (except for the replicates). This yielded the following experimental run programme as shown in Table 2.

TABLE 2

Experimental runs

Day No.	Treatment combination (randomly selected)	Treatment	Response Variable*
1	B*l* P*l* S*h* R*l*	1	33,6
2	B*l* P*l* S*h* R*l*	1	34,2
3	B*l* P*l* S*l* R*l*	2	35,1
4	B*h* P*l* S*h* R*l*	3	32,8
7	B*h* P*l* S*h* R*h*	4	32,2
8	B*h* P*h* S*l* R*l*	5	28,5
9	B*h* P*l* S*l* R*h*	6	29,3
10	B*h* P*l* S*l* R*l*	7	30,1
11	B*l* P*l* S*h* R*l*	1	33,7
14	B*l* P*l* S*h* R*h*	8	31,1
15	B*h* P*h* S*h* R*l*	9	34,3
16	B*l* P*h* S*h* R*l*	10	34,2
17	B*l* P*h* S*l* R*h*	11	34,2
18	B*l* P*l* S*h* R*l*	1	34,6
21	B*l* P*h* S*h* R*h*	12	33,5
22	B*l* P*l* S*l* R*h*	13	32,1
23	B*h* P*h* S*h* R*h*	14	30,6
24	B*h* P*h* S*l* R*h*	15	34,0
25	B*l* P*h* S*l* R*l*	16	35,0
28	B*l* P*l* S*h* R*l*	1	34,8

* The response variable selected as a measure of massecuite exhaustion was the apparent purity of the final nutsch molasses which was tested. The lower the apparent purity the greater the exhaustion.

Analysis of the results

A computer is useful for making the repetitive type of calculations that are required[8]. However, it is possible to extract information on the main and interactive effects by simple methods.

The effect of massecuite brix on final molasses apparent purity may be illustrated as shown in Table 3.

TABLE 3

Molasses apparent purities for the various combinations at high and low massecuite brix

Levels of the other Factors	Molasses Apparent Purity B*l*	B*h*	B*h* – B*l*
P*l* S*l* R*l* . . .	35,1	30,1	– 5,0
P*l* S*l* R*h* . . .	32,1	29,3	– 2,8
P*l* S*h* R*l* . . .	34,2	32,8	– 1,4
P*l* S*h* R*h* . . .	31,1	32,2*	1,1
P*h* S*l* R*l* . . .	35,0	28,5	– 6,5
P*h* S*l* R*h* . . .	34,2	34,0	– 0,2
P*h* S*h* R*l* . . .	34,2	34,3*	0,1
P*h* S*h* R*h* . . .	33,5	30,6	– 2,9
		Average:	– 2,2

In all except two cases (indicated by an asterisk), the high massecuite brix results in lower final molasses apparent purity. This effect may be quantified by simply averaging the observed differences in apparent purity between the high and low levels of brix which gives an overall average brix effect of –2,2 units. This indicates that, on average, the effect of changing brix from its low to its high level is to decrease the molasses apparent purity by 2,2 units. The other main effects may be investigated in the same way.

The interactive effects may be studied by selecting the massecuite brix/stirring rate interaction for example; this may be illustrated by the data shown in Table 4.

TABLE 4

Molasses apparent purities for the brix/stirring interaction

Purity/Retention Levels	Brix/Stirring Levels B*l* S*h*	B*h* S*l*	B*l* S*l*	B*h* S*h*
P*l* R*l*	34,2	30,1	35,1	32,8
P*l* R*h*	31,1	29,3	32,1	32,2
P*h* R*l*	34,2	28,5	35,0	34,3
P*h* R*h*	33,5	34,0	34,2	30,6
Column average . .	33,3	30,5	34,1	32,5
Average	31,86		33,29	

If there is no interaction the effect of changing the level of one factor will be independent of the level of the other and thus the change in average molasses purity between the B*l* S*h* and B*h* S*h* levels and between the B*l* S*l* and B*h* S*l* levels will be the same.

The interactive effect between brix and stirring rate is 1,4 units, which is determined by differencing the average purity of the B*l* S*l*, B*h* S*h* columns (33,29) and the B*l* S*h*, B*h* S*l* columns (31,86). One way to visualise the effect of this interaction is to use the diagram depicted in Figure 2.

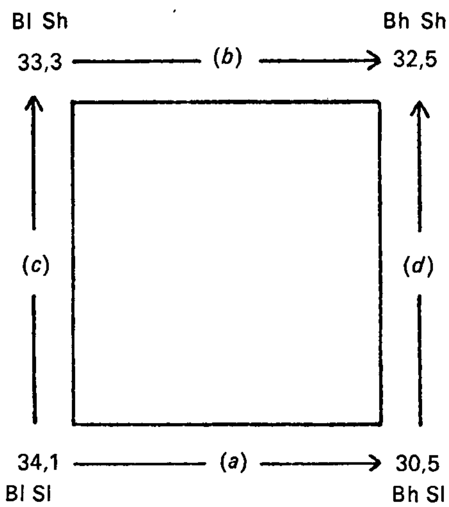

FIGURE 2 Brix / Stirring Interaction.

The figures at the corners of the square are the column averages shown in Table 4.

At low (S*l*) and high stirring (S*h*) increasing the massecuite brix reduces molasses apparent purity (as indicated by case (*a*) and (*b*) respectively). At low brix (B*l*) an increase in stirring decreases molasses purity (case (*c*)), and at high brix (B*h*), an increase in stirring increases molasses purity (case (*d*)).

In practical terms this means that high brix should be the first aim. If this is achieved, then high stirring should not be used. If high brix is however not possible, then high stirring becomes beneficial.

This interaction could be explained by heat generation in the high brix/high stirring case. Heat generated would not be capable of affecting the B*l* S*l* to B*l* S*h* and B*l* S*h* to B*h* S*h* transitions but would affect the B*h* S*l* to B*h* S*h* one. The other second order interactions may be similarly investigated. Third and higher order interactions are, however, more difficult to interpret using graphical methods.

Main and interactive effects cannot, however, be judged in the proper perspective unless something is known of the intrinsic variability of the procedure. To quantify the uncertainty surrounding these calculated effects the experimental error is estimated by determining the variance of replicate observations, and calculating confidence intervals for each effect.

The 99 % confidence intervals for the main effects and second order interactions are shown in Figure 3. The method of calculation and meanings of the various terms are discussed in Appendix 1.

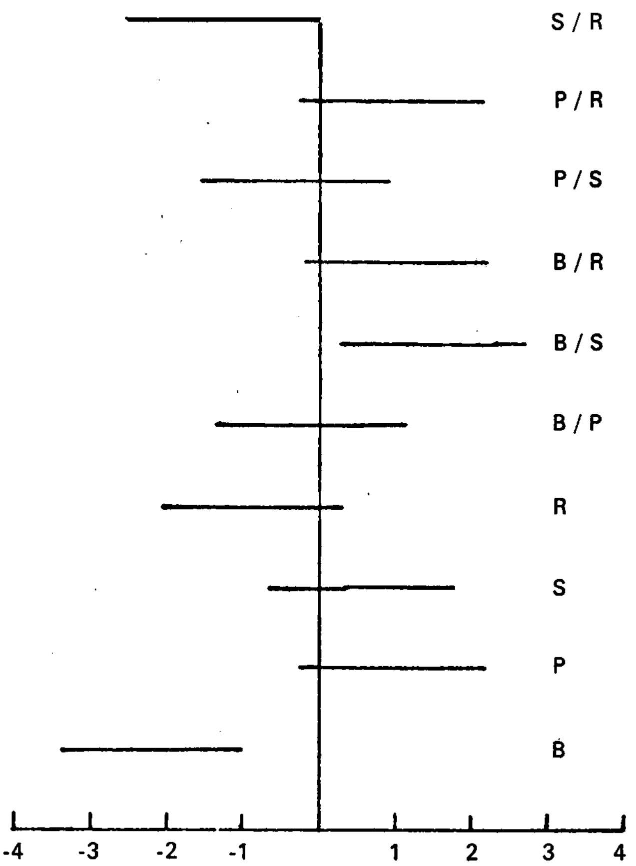

FIGURE 3 99% Confidence Intervals for main and second order interactions.

Thus, massecuite brix at pan drop is the most important factor and this finding is statistically significant at the 1% level which means that there is one possibility in 100 that the finding is due to chance and thus not genuine.

The 2 interactions, stirring/retention and brix/stirring, are also significant at the 1% level and this information is used to study these effects as described previously. If this is done it is found that the stirring/retention interaction yields results which cannot be valid from a technological point of view. This illustrates a point which needs to be stressed — statistics need more than significance and other tests for their proper interpretation. It is inherent in the nature of experimentation that strange results will appear from time to time. This particular interaction is thus rejected, although indicated to be significant at a relatively high level.

Acknowledgements

The authors' thanks are due to the Board of Directors of Huletts Sugar Limited for permission to publish this paper.

REFERENCES

1. Davies, Owen L. (Ed.) (1963). The Design and Analysis of Industrial Experiments, Oliver and Boyd, Edinburgh.
2. Deenick, Z. McMartin, A. and Colepeper, J. (1939). A factorial experiment with fertilisers on sugar cane. SASTA Proc **13,** p 81-88.
3. Hahn, J. G. (1975). Designing experiments, Parts 1 and 2. Chemtech Aug/Sept p 496-498; p 561-562.
4. Hill, W. J. and Demler, W. R. (1970). More on planning experiments to increase research efficiency. Ind.Eng.Chem. **62,** No. 12, p 60-65.
5. Hunter, W. G. and Hoff, M. E. (1967). Planning experiments to increase research efficiency. Ind.Eng.Chem., **59,** No. 3, p 43-48.
6. Juran, J. M. (Ed.) (1974). Quality Control Handbook, McGraw-Hill, Section 27.
7. Pavelic, V. and Saxena, U. (1969). Basics of statistical experiment design. Chem. Eng. Oct. 6, p 175-180.
8. Ralson, A. and Wilf, H. (Eds.) (1962). Mathematical Methods for Digital Computers. John Wiley & Sons, p 221-230.
9. Stowe, A. R. and Mayer, R. P. (1966). Efficient screening of process variables. Ind. Eng. Chem., **58,** No. 2, p 36-40.
10. Szonyi, G. (1973). Experimental design and optimisation the easy way. Part I: Factorial design. Chemtech, Jan., p 36-44.

Appendix 1
Calculation of confidence intervals

The equation for the 99% confidence interval for each effect is:

$$\text{Confidence Interval} = \text{Effect} \pm t_{\varnothing,\ \alpha/2} \times s/2$$

Where Effect is the numeric value calculated for each effect.

$\varnothing = n - 1$ where n is the number of replicate observations.

α is the significance level = 0,01 in the case of 99% confidence intervals.

$t_{\varnothing,\ \alpha/2}$ is the value read from a table of probability points of the t-distribution.

s is the standard deviation of the replicate observations and is calculated by

$$s = \sqrt{1/\varnothing \sum_{i}^{n} (y_i - \bar{y})^2} = 0{,}53$$

Thus the 99% confidence interval for the brix effect is given by

$$- 2{,}2 \pm t_{4;0{,}01} \times 0{,}27$$
$$= + 2{,}2 \pm 4{,}6 \times 0{,}27$$
$$= (- 3{,}44 - 0{,}96)$$

As this interval does not include the origin it is said to be significant at the 1% level.