The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

A COMPARISON OF RESPONSE SURFACE AND FACTORIAL DESIGNS IN AGRICULTURAL RESEARCH

R. J. Williams* and J. R. Baker†

The use of response surface designs in preference to factorial designs in agriculture has become widely discussed and recommended. This paper simulates data for comparable response surface and factorial designs and uses this to demonstrate the similarities between the designs and their analyses and at the same time to point out some of the customary differences in their analyses. More particularly, it aims to show: (a) that for any design a reduction in the number of plots sown reduces the reliability of the results, and (b) that both a response surface design and a factorial design allow the calculation of estimated equations of response surfaces.

INTRODUCTION

Since the publication by Box and Wilson[1] of their first study of response surface designs the interest in the subject has been such that 125 papers are listed in a recent bibliography by Hill and Hunter[2] of theoretical and applied aspects of these designs. The earlier designs were conceived for experimentation in sciences such as physical chemistry where experimental error was small, the region of optimum treatment combination was unknown and conditions were sufficiently stable over a period of time to allow for sequential experiments. In agriculture these conditions rarely apply; but despite this, the use of response surface designs in Australian agriculture has become more common and their use discussed and recommended by a number of economists[3]. Such recommendations are usually based on two concepts: (a) that because response surface designs test a greater number of distinct levels for each independent variable (factor), they require little or no replication and hence allow a reduction in the number of plots sown, and (b) that response surface designs lend themselves to the calculation of a second degree prediction equation which when graphed gives a reasonable representation of the true response surface.

---

* Statistician, Philip Shrapnel and Co. Ltd.

† Research Biometrician, Australian Fertilizers Ltd.

[1] G. E. P. Box and K. B. Wilson, "On the Experimental Attainment of Optimum Conditions". *Journal of the Royal Statistical Society*, vol. 13, no. 1, 1951, pp. 1–38.

[2] W. J. Hill, and W. G. Hunter, "A Review of Response Surface Methodology: A Literature Survey". *Technometrics*, vol. 8, no. 4 (November, 1966), pp. 571–90.

[3] For example, J. L. Dillon "Economic Considerations in the Design and Analysis of Agricultural Experiments", this *Review*, vol. 34, no. 2 (June, 1966), pp. 64–75.

Implicit in the recommendation of response surface designs is often an almost complete condemnation of "comparable" factorial designs. Box and Hunter[4] mathematically defined comparable or competing designs. Here it suffices to say that comparable designs contain the same number of factors and span approximately the same experimental region and that the response surface designs and factorial designs used in the present paper are comparable according to the Box and Hunter definition.

What is sometimes overlooked in claiming the superiority of response surface designs relative to factorial designs is that a response surface equation can be estimated for factorial designs as well as response surface designs. Regardless of the design the statistical analysis is done by least squares multiple regression techniques. Perhaps these techniques are disguised somewhat by resorting in one or more of the analyses to simple text-book procedures, still it remains that they are derived initially by these regression analyses. So it should not be surprising when experiments using comparable designs are analysed by these very similar methods to determine an approximation to the same "true" response surface that the results are virtually the same.

The general aim of this paper is to demonstrate the similarities between competing response surface and factorial designs in their design and analysis, and at the same time to point out some of the customary differences in their analysis and in the interpretation of results from them. More particularly, and relative to concepts (a) and (b) above, it aims to show[5] that for any design a reduction in the number of plots sown reduces the reliability of the results, and that (as indicated by Robinson and Nielson[6]), both a response surface and a factorial design allow the calculation of estimated equations of response surfaces. This latter point is further laboured by the inclusion of a "random"[7] design and the estimation of a response surface for it.

METHOD

The most obvious way of comparing designs would be to establish experiments in adjoining areas using the different designs so that for each a set of yields was available. This procedure, however, would not resolve the difficulty since it provides no basis for the selection of a particular design after the inspection of the manipulation of the sets of yields. Even if such a basis were available it would be necessary to repeat the experiments a number of times to establish their consistency because of the relatively high variability of agricultural experiments. The cost would be prohibitive.

---

[4] G. E. P. Box and J. S. Hunter, in V. Chew, *ed.*, *Experimental Designs in Industry* (New York: Wiley, 1958.)

[5] "Show" is used in the colloquial sense, i.e. it implies a demonstration. Obviously, with a comparison of only one type of each design, no mathematical proof is contemplated.

[6] P. Robinson and K. F. Nielson "Composite Designs in Agricultural Research", *Canadian Journal of Soil Science*, vol. 40, no. 2 (August, 1960), pp. 168–76.

[7] This is not a completely random design—it is a hybrid of the response surface and factorial designs.

The statistical comparison of designs for response surface estimation is a fairly complex technique, but it does include a number of criteria for discrimination which are the more useful because of their limited dependence on the yields observed in experiments. These criteria are concerned with the biases and variances of the coefficients in the prediction equation, the accuracy of the estimate of experimental error, and the information distribution. Collectively they remove much of the subjective element from the choice of the design.

The procedure used for this paper was to generate a surface of "true" yields (Z, "dependent" variables) by a second degree equation in two independent variables X and Y corresponding to two different fertilizers. Yields for a number of fertilizer combinations (points) appropriate to the designs to be compared were simulated by using the equation to calculate the yield corresponding to each point and adding to this, "error" terms and constants corresponding to a replication effect. The equation used for the surface was

$$Z = -48{\cdot}16 + 2{\cdot}57X + 63{\cdot}62Y - 0{\cdot}0181X^2 - 12{\cdot}900Y^2 - 0{\cdot}2022XY$$

which is the equation found by Box and Hunter[8]. Its use was considered appropriate since within the experimental region it approximates well a typical second degree fertilizer response surface, but characteristically beyond that region it is of no value in prediction. The error terms were selected at random from a normal population with mean zero and variance 36, thus giving a coefficient of variation of about 8 per cent. This is typical of wheat fertilizer experiments in N.S.W.

Three designs were considered, each "centred" at the point $X = 45$, $Y = 2$:

(a) A response surface design as used by Box and Hunter[8]. This was a second order rotatable hexagonal with six peripheral and four central points[9].

(b) A 3×3 factorial design.

(c) A design in which 30 points were chosen at random from the points used in designs (a) and (b). Three replications of each of designs (a) and (b) were used and statistical analysis of yield was carried out using, firstly the response surface approach, and secondly the factorial approach. The experimental region for each is shown diagrammatically in figure 1.

In the discussion of results from these designs, emphasis is placed on the statistical criteria for comparison of designs.

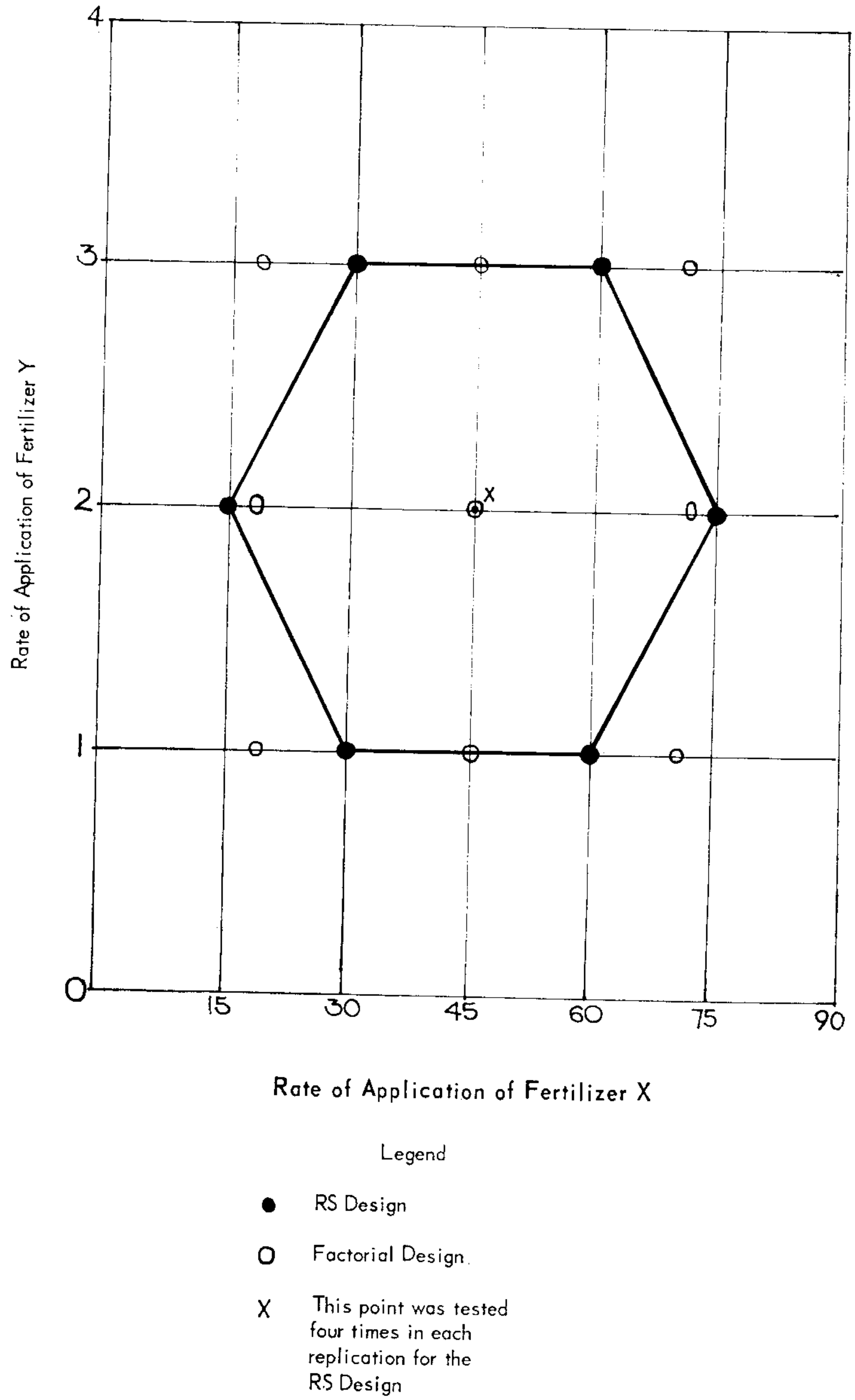
---

[8] Box and Hunter, *op. cit.*

[9] The four central points are necessary for rotatability but in the context of Box and Hunter (*op. cit.*) also enable an estimate of variance. However, as agricultural experiments are not normally considered without allowing some internal estimate of error variance the replicated central points are not used for this purpose.

FIGURE 1

Distribution of Experimental Points



COMPARISON OF STATISTICAL ANALYSES

Response surface analyses usually present the estimates of the regression coefficients (b_i, b_{ij}), an indication of their significance (by a t-test comparison with their standard errors) and an analysis of variance. Also a lack-of-fit mean square is given to provide an indication of whether the second degree model is adequate; if this mean square is significantly greater than error mean square then perhaps an equation involving terms of higher degree, e.g. in X^3, X^2Y etc., might be necessary.

Factorial analyses, on the other hand, present an analysis of variance with the treatment mean square in the form of individual components corresponding to linear (L), quadratic (Q), $L \times L$, $L \times Q$, etc. mean square, each with a single degree of freedom. These components are readily calculated and, when used in conjunction with a two-way table of means and appropriate differences for significance, are of considerable assistance in the interpretation of experimental results and in the making of recommendations for future fertilizer application rates.

The mean squares for X, X^2, XY etc. in the response surface analysis must be considered as quite distinct from the mean squares for X-linear, X-quadratic, X-linear $\times$ Y-quadratic interaction etc. in the factorial analysis. Only a few of these components are equal in pairs for designs where the factors involved are repeated equally at each equally-spaced level (as in factorials).

Despite the differences in the presentation of the two types of analysis of variance, they have much in common. Table 1 gives the analysis of variance for each of the three designs analysed as far as is thought sensible by each of the two methods. These analyses of variance have been arranged to emphasize the fact that for a set of yields corresponding to a given design the mean squares for replications, for treatments, and for error take values which are independent of the type of analysis used. It follows that the standard deviations, coefficients of variation, and coefficients of determination (R^2) are equal for the two types of analyses. These equalities are sometimes obscured by allocating certain components of treatment mean square to error mean square, usually with the purpose of obtaining more degrees of freedom for the latter. For example, with response surface analyses a non-significant lack-of-fit mean square may be included in error mean square, while for factorial analyses, non-significant components, e.g. $L \times Q$, $Q \times Q$, of the two-factor interaction may be re-allocated.

Both types of statistical analysis aim to produce an optimum rate of fertilizer application in the future. However, the approaches of each are very different. Agricultural scientists using the response surface analysis employ the analysis of variance and its associated lack-of-fit mean square mainly to ensure that the statistical model proposed adequately represents the data, and, having been reassured on that count, make objective predictions using a prediction equation based on the regression coefficients. Little attention is paid to mean yields for individual treatment combinations. Those using the factorial analysis,

Table 1

Analyses of Variance—Mean Squares

Response Surface Design						Factorial Design						"Random" Design		
Response Surface-type Analysis			Factorial-type Analysis			Response Surface-type Analysis			Factorial-type Analysis					
Source	*df*	*MS*	Source	*df*	*MS*	Source	*df*	*MS*	Source	*df*	*MS*	Source	*df*	*MS*
Replications	2	3·100	*Replications*	2	3·100	*Replications*	2	51·593	*Replications*	2	51·593			
Treatments	6	711·550	*Treatments*	6	711·550	*Treatments*	8	789·342	*Treatments*	8	789·342	*Treatments*	11	564·552
X	1	2146·758	X (ignoring Y)	4	985·572	X	1	3472·222	X-Linear	1	3742·222			
Y	1	27·000	Y (ignoring X)	2	440·345	Y	1	200·000	X-Quadratic	1	1390·297			
									Y-Linear	1	200·000			
Additional for X^2, Y^2, XY	3	698·040				Additional for X^2, Y^2, XY	3	869·485	Y-Quadratic	1	1157·407			
Lack-of-fit	1	1·422				Lack-of-fit	3	11·377	Lin-Lin	1	60·750			
									Curvatures	3	11·355			
Residual	21	40·227	*Residual*	21	40·227	*Residual*	16	35·551	*Residual*	16	35·551	*Residual*	16	33·803
Total	29		Total	29		Total	26		Total	26		Total	29	

Response Surface Design:
Standard Deviation 6.50
Coefficient of Variation: 8·3%
Coefficient of Determination: 82·8%

Factorial Design:
Standard Deviation: 5·96
Coefficient of Variation: 8·2%
Coefficient of Determination: 92·0%

"Random" Design:
Standard Deviation: 5·81
Coefficient of Variation: 8·3%
Coefficient of Determina ion: 91·8%

on the other hand, make their prediction somewhat subjectively on the mean treatment yields presented in a two-way table, the statistically significant mean squares in the analysis of variance being of prime importance in the interpretation of that table. Here regression coefficients are not calculated and a prediction equation is not formed.

COMPARISON OF STATISTICAL DESIGNS

Regression techniques for the calculation of a prediction equation for a response surface are not restricted to response surface designs but can be applied to any design with quantitative levels. The sole reason for the inclusion of a random design in this paper is to emphasize this fact. Table 2 gives the regression coefficients for each of the designs and for the "true" coefficients. The coefficients for the random design are compatible with those for the two other designs and those for the "true" surface. However its use in fertilizer experiments will generally be restricted to situations where many missing plots or sowing mistakes make a designed experiment non-orthogonal, and where an electronic computer is available to handle the considerable calculations involved. No further mention of the random design will be made in this paper.

For the more pertinent comparison between the response surface design and the factorial design, the analysis of variance of table 1 show that, as expected, the treatment and error mean squares are quite similar for the two designs. For the response surface analysis, the lack-of-fit mean square is not significant in either case, indicating the adequacy of the quadratic model for each design. No importance should be attached to the fact that the coefficient of determination (R^2) is greater for the factorial design (92 per cent) than for the response surface design (82 per cent) apparently indicating that the response surface for the former is a better fit than that for the latter. The difference occurs because in simulating the data, errors were allocated at random: if a series of further comparisons of R^2 were made the average values of R^2 for the two designs would be nearly equal.

Similar variation due to random allocation of errors is shown in the regression coefficients in Table 2. There the coefficients for the two designs are again compatible with each other and with the "true" values but the differences between a few pairs of coefficients seem quite large, perhaps suggesting a material difference in the results from the two designs. However the measure of similarity of the two results is not provided by comparison of the individual coefficients, but rather by the correlation between yields predicted from the whole set of coefficients. This measure is made by calculating predicted yields for the response surface design, the factorial design, and the "true" equations, for a grid of 15 points in the experimental region. The correlation coefficients for these yields are 0·95 for response surface design and factorial design yields, 0·98 for response surface design and "true" yields and 0·99 for factorial design and "true" yields. Collectively these results indicate that the three surfaces are very similar in the experimental region.

Table 2

Partial regression coefficients (± standard errors) for response surface analysis: All replications included in analysis

	b_0	b_1	b_2	b_{11}	b_{22}	b_{12}
"True" value	—48·16	2·57	63·62	—0·018	—12·900	—0·202
Response surface estimate ..	—51·17	2·71 ± 0·41	65·17 ± 11·16	—0·017 ± 0·004	—12·167 ± 2·382	—0·333 ± 0·122
Factorial estimate	—52·72	2·93 ± 0·36	62·94 ± 9·74	—0·024 ± 0·004	—13·889 ± 2·300	—0·090 ± 0·065
"Random" estimate	—42·88	2·50 ± 0·33	61·93 ± 9·55	—0·018 ± 0·003	—13·760 ± 2·243	—0·144 ± 0·068

The magnitude of bias in the regression coefficients is one of the statistical criteria for comparison of designs. Such bias will occur in some of the coefficients if the model used does not adequately fit the data, as would be the case if a quadratic equation was used when a cubic was needed. The non-significant lack-of-fit mean square for both the response surface design and factorial design provides a satisfactory indication that any biases in the coefficients are small. However, had the quadratic model been inadequate, the method of Box and Hunter[10] indicates that the biases for the response surface design would have been marginally greater than those for this particular factorial design. This type of theoretical calculation of potential bias assumes importance when it is realized that the lack-of-fit test of actual bias can be made only after the experiment is completed. A significant bias found at this stage can only be materially reduced by the use of a higher order model and this might not be possible because of a lack of sufficient experimental points. The likelihood and consequences of these biases should be evaluated at the time of choosing a design.

REDUCTION IN PLOT NUMBERS

For response surface analyses, further statistical criteria for comparison of designs involve the comparison of the variances of the regression coefficients and the assessment of the accuracy of the estimates of experimental error. The variances of the regression coefficients are much the same for the response surface design and factorial design and for these designs unbiased estimates of error mean square based on approximately the same number of degrees of freedom are made. Accordingly the statistical criteria given above will not be used to contrast these designs: instead they will be used to examine the consequences of a reduction in plot numbers for each design. A reduction can be effected by omitting some treatment combinations, thereby changing the design, or by reducing the number of replications. This latter procedure is the one demonstrated here.

The variances of the regression coefficients are functions of the number of plots in an experiment and of the experimental error. For a given estimate of variance a reduction from 3 to 2 or from 3 to 1 replications increases the variance of the regression coefficients by a factor of 1·5 or 3 respectively for both designs. Similarly estimates of the error mean square become less reliable as plot numbers are reduced, again regardless of design. The combination of these two effects results in increasingly erratic estimates of the coefficients with decreases in plot numbers.

A practical demonstration of these theoretical considerations is given in tables 2, 3, and 4. The latter two tables give the estimates and their standard errors for each design when first one replication and then two replications are omitted from the experiment. As expected, for each design the estimates range more widely about the "true" value as the number of replications is reduced, and equivalently their standard errors are larger. That this should be so becomes more obvious when it is

[10] Box and Hunter, *op. cit.*

Table 3

Partial regression coefficients (± standard errors) for response surface analysis: Two replications included in analyses

Design	b_0	b_1	b_2	b_{11}	b_{22}	b_{12}	Variance
"True" value ..	—48·16	2·57	63·62	—0·018	—12·900	—0·202	36·00
Response Surface Reps 1, 2 ..	—44·38	2·61 ± 0·54	61·50 ± 14·92	—0·015 ± 0·005	—11·375 ± 3·189	—0·350 ± 0·163	48·22
Response surface Reps 2, 3 ..	—60·94	3·28 ± 0·47	61·25 ± 12·91	—0·023 ± 0·004	—10·562 ± 2·752	—0·375 ± 0·142	35·91
Response surface Reps 1, 3 ..	—48·19	2·25 ± 0·52	72·75 ± 14·33	—0·014 ± 0·005	—14·563 ± 3·051	—0·275 ± 0·157	44·14
Factorial Reps 1, 2	—50·34	3·09 ± 0·47	55·37 ± 12·70	—0·025 ± 0·005	—11·583 ± 3·000	—0·110 ± 0·085	35·92
Factorial Reps 2, 3	—52·53	2·93 ± 0·36	65·23 ± 9·85	—0·025 ± 0·004	—14·750 ± 2·325	—0·005 ± 0·066	21·63
Factorial Reps 1, 3	—53·28	2·76 ± 0·48	68·23 ± 12·97	—0·022 ± 0·005	— 15·333 ± 3·062	—0·105 ± 0·087	37·51

Table 4

Partial regression coefficients (± standard errors) for response surface analysis: One replication included in analyses

Design	b_0	b_1	b_2	b_{11}	b_{22}	b_{12}	Variance
Response surface Rep 1	—31·63	1·58 ± 0·72	73·00 ± 19·75	—0·006 ± 0·006	—15·375 ± 4·218	—0·250 ± 0·216	42·17
Response surface Rep 2	—57·13	3·64 ± 0·76	50·00 ± 20·69	—0·024 ± 0·007	— 7·375 ± 4·419	—0·450 ± 0·227	46·29
Response surface Rep 3	—64·75	2·92 ± 0·63	72·50 ± 17·25	—0·021 ± 0·005	—13·750 ± 3·683	—0·300 ± 0·189	32·17
Factorial Rep 1 ..	—49·00	2·91 ± 1·13	58·37 ± 30·67	—0·023 ± 0·012	—12·167 ± 7·239	—0·160 ± 0·205	104·81
Factorial Rep 2 ..	—51·59	3·27 ± 0·46	52·37 ± 12·50	—0·028 ± 0·005	—11·000 ± 2·953	—0·060 ± 0·084	17·44
Factorial Rep 3 ..	—57·47	2·60 ± 0·56	78·08 ± 15·32	—0·022 ± 0·006	—18·500 ± 3·620	—0·050 ± 0·102	26·14

realized that the estimates in table 2 are the averages of those in tables 3 and 4.

The trend to instability in error mean square is most noticeable for the factorial design in table 4. This is hardly surprising since the mean square is based on only three degrees of freedom.

Conclusions identical to those above will be reached when any one experiment is compared with another with fewer plots. The smaller designs may or may not be slightly more efficient on a per plot basis but even a very small reduction in plot numbers negates this advantage. Box[11] reports success in a number of multi-factor experiments with response surface designs using up to twenty plots. These were all conducted in the field of physical chemistry where the coefficient of variation is of the order of 1 per cent. In Eastern Australian agriculture, examples of coefficients of variation range from 5 to 8 per cent for experiments with pigs, 7 to 15 per cent for grain crops, 15 to 25 per cent for temperate pasture, and up to 50 per cent for sub-tropical legume pasture mixtures. In this context any experiment of either design using less than twenty plots will provide predictions of yield with such high variability that basing practical development programmes on them is scarcely feasible. Agricultural scientists and agricultural economists with resources thus limited should consider whether derived predictions would be of greater or of less reliability than the opinions of an informed farmer or agronomist.

DISTRIBUTION OF INFORMATION

Statistically speaking, "information" at a point is defined as the reciprocal of the variance of an estimate at that point. Intuitively speaking, it gives a numerical measure of the reliability of an estimate at a point. In the response surface design used throughout this paper the centre point is duplicated four times in each replication so that the information near this point is much greater than at the same point for the factorial design. On the other hand the information near each "corner" of the factorial design is greater than at identical points for a response surface design. Further, for a response surface design with more levels of a factor in a given range than a factorial design, the information will be more evenly distributed along the line of that factor than for the factorial design, where there will be areas of very high information near the experimental points and of lower information between them.

Thus in the design of an experiment, after the experimental region has been determined and the possible number of plots approximated, the final decision between competing designs should include consideration of information contour diagrams for each which indicate those sub-regions in which information is high. Subsequently a design should be chosen and placed so that regions of high information coincide with the suspected areas of maximum interest to the scientist.

---

[11]G. E. P. Box, "The Exploration and Exploitation of Response Surfaces: Some General Considerations and Examples", *Biometrics*, vol. 10, no. 1 (March, 1954), pp. 16–60.

CONCLUSIONS

We consider that the development of response surface designs has been important in agriculture not for the designs themselves but because of the type of analysis associated with them. The mathematical definition of continuous surfaces has added a most useful statistical technique to those already in existence. However, this should be seen only as an addition; the best statistical analysis and the most reliable predictions will make use of as many techniques as are sensibly available. In this respect, factorial designs have an advantage over response surface designs in that they allow a full response surface analysis in conjunction with the usual factorial analysis. Response surface designs are not flexible in this sense, and their use in agriculture is limited accordingly.

Obviously we do not classify fractional replicates of factorial designs as "response surface designs", since it is common knowledge that these fractional replicates are very useful in agricultural experiments with large numbers of factors. We take this attitude because it has been only in the last ten years that these designs, many of which have been available for thirty years, have been frequently classified as response surface designs, and because a response surface equation based on them has rarely been calculated in agriculture.

With regard to central composite designs where a number of extra points including central points are added to a full or fractional factorial design, it should be realized that fertilizer response experiments are subject to much uncontrolled variation in environment caused by weather and soil variations. Accordingly the optimum fertilizer combination, however defined, will vary considerably from year to year and place to place. Any design which concentrates much of the information in an experiment in a small sub-region in the hope that this will be the region of the optimum, and which has a low distribution of information elsewhere, will often provide little information in the actual optimum region. For this reason most central composite designs are unsuitable for fertilizer response experiments.

The effective use of response surface designs in agriculture would appear to be restricted to rather unusual experiments such as where a central composite design can be used, or where the experimental region is an unusual shape or where designs with special patterns of information distribution are required. In the absence of these special conditions it is suggested that factorial designs be used and that the resultant yields be analysed using both the factorial and response surface analysis.