Optimum and Other Response Surface Designs. Comments on "Response Surface Design Evaluation and Comparison" by Anderson-Cook, Borror and Montgomery

Anthony C. Atkinson,* Department of Statistics,
London School of Economics, London WC2A 2AE, UK

1 Introduction

It is a pleasure to comment on this paper that contains a wealth of recent references on the assessment and comparison of response surface designs. Despite the length of my contribution, space precludes me from commenting on many interesting aspects.

One purpose of these comments is to provide references to related work, particularly on optimum designs. I establish the customary notation for optimum designs in §2. Section 3 contains comments on a number of points raised by the authors. In §4 I give a few small response surface designs with high symmetry, some of which are new, that experimenters may find useful. Maybe the authors would like to evaluate these designs.

2 Optimum Experimental Design

2.1 Exact and Continuous Designs

The main tools that the authors use are based on the properties of the standardized prediction variance, defined by the authors in their §2.1. The importance of this quantity comes, as the authors say, from the theory of optimum design, with the target maximum value determined by the equivalence between continuous D-

*a.c.atkinson@lse.ac.uk

and G-optimum designs. To relate their work more closely to parallel work on optimum design, it is helpful to establish the standard notation of that literature.

Consider an experiment in which there are m factors and write the linear model as

$$\mathrm{E}(y) = F\beta.$$

Here y is the $N \times 1$ vector of responses, β is a vector of p unknown parameters to be estimated by least squares and F is the $N \times p$ *extended* design matrix. The ith row of F is $f^{\mathrm{T}}(x_i)$, a known function of the m explanatory variables that can take values in the design region $\mathcal{X}$. For the second-order model with $m = 2$

$$\mathrm{E}(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2, \tag{1}$$

and

$$f^{\mathrm{T}}(x_i) = (1 \quad x_{1i} \quad x_{2i} \quad x_{1i}^2 \quad x_{2i}^2 \quad x_{1i}x_{2i}).$$

This notation can be extended to compound design criteria, §3.5, in which the jth model has extended design matrix F_j.

The design is determined by the experimental values of x and by the numbers of replications n_i or, equivalently, the weights $w_i = n_i/N$ at each x_i. The design can be written as a measure ξ that puts weight w_i at x_i, that is

$$\xi = \left\{ \begin{array}{cc} x_1 & x_2 \ldots x_n \\ w_1 & w_2 \ldots w_n \end{array} \right\},$$

with ξ_N the exact design in which $w_i = n_i/N$. The theory was developed by Kiefer whose work on optimum design is collected in Brown et al. (1985). Use of the continuous design, to which the equivalence theory applies, removes dependence of the design on N. Any practical design for N trials will be exact with integer replication n_i at each design point. Exact and continuous D-optimum designs may be identical, but they usually are not in response surface work.

The least squares estimator of the parameters is

$$\hat{\beta} = (F^{\mathrm{T}}F)^{-1}F^{\mathrm{T}}y,$$

where the $p \times p$ matrix $F^{\mathrm{T}}F$ is the information matrix for β. For the design measure ξ we consider instead the standardised information matrix

$$M(\xi) = F^{\mathrm{T}}WF, \tag{2}$$

where W is a diagonal matrix with elements w_i.

The variance of the predicted response $\hat{y}(x)$ for an exact design is conveniently written as

$$d(x, \xi_N) = \frac{N\mathrm{var}\{\hat{y}(x)\}}{\sigma^2} = f^{\mathrm{T}}(x)M^{-1}(\xi_N)f(x), \tag{3}$$

the quantity plotted for several designs in the authors' Figure 1. For the continuous design (3) becomes

$$d(x, \xi) = f^{\mathrm{T}}(x) M^{-1}(\xi) f(x).$$

2.2 The General Equivalence Theorem and Design Efficiency

Exact D-optimum designs maximize $|M(\xi_N)|$ and exact G-optimum designs minimize the maximum over $\mathcal{X}$ of $d(x, \xi_N)$ given by (3). We write this maximum for some ξ as

$$\bar{d}(\xi) = \max_{x \in \mathcal{X}} d(x, \xi).$$

The equivalence theorem for D- and G-optimality holds for continuous designs. If ξ^* is a continuous D-optimum design, then $\bar{d}(\xi^*) = p$, the number of parameters of the linear model. Exact D-optimum designs may have values of $\bar{d}(\xi_N^*)$ slightly larger than p.

The D-efficiency of a design ξ is

$$D_{\text{eff}} = \left\{ \frac{|M(\xi)|}{|M(\xi^*)|} \right\}^{1/p}. \tag{4}$$

The comparison of information matrices for designs that are measures removes the effect of the number of observations, while taking the pth root of the ratio of the determinants in (4) results in an efficiency measure which has the dimensions of a variance, irrespective of the dimension of the model.

The G-efficiency of the design ξ is

$$\mathrm{G}_{\text{eff}} = \bar{d}(\xi^*)/\bar{d}(\xi) = p/\bar{d}(\xi).$$

A recent book-length treatment of optimum experimental design, including both theory and SAS code for the construction of designs, is Atkinson, Donev, and Tobias (2007) which also provides numerous references to the literature on optimum design.

3 Detailed Comments

3.1 Graphical Displays

The distinction between exact and continuous optimum designs is nicely made by two-factor designs for a circular region as illustrated in the variance-dispersion graph of the authors' Figure 1.

D-optimum response surface designs are described by Farrell, Kiefer, and Walbran (1968) for three types of design region. For spherical regions the D-optimum design puts a weight $2/\{(m+1)(m+2)\}$ at the centre point, with the rest of the weight distributed uniformly around the circumference of the design region. For $m = 2$ the region is a circle and the weight at the centre is 1/6. Thus one continuous and exact D-optimum design consists of points in a regular pentagon with one at the centre. In Figure 1 the authors have plotted the properties of a hexagonal design. This has weight 1/7 at the centre and so is not quite the continuous D-optimum design. This is shown by the value of $d(x, \xi)$ being slightly more than 6 at the centre of the region and slightly lower than 6 at the edge.

As the authors comment, all three designs in Figure 1 are rotatable. If they are not, it is informative to consider the minimum and maximum variance at each radius, as in the original plots of Box and Behnken (1960), and perhaps the average variance as well. It is also necessary to decide how to treat designs over a cubical region. For example, should the calculations for a radius greater than one include only those parts of $\mathcal{X}$ for that radius or all points on the circle or sphere?

The fraction of design space plot overcomes these problems. However, there are a few outstanding questions. One, mentioned by the authors at the end of §3, is that the plots are for measures, so information on the number of trials is lost when designs for different N are being compared. I discuss design optimality and costs in §3.3.

A final point about these useful graphical tools is that, on both Figure 1 and Figure 2 it has been felt necessary to include lines of efficiency. An alternative would be to plot efficiency rather than variance.

3.2 The Comparison of Small Designs

Towards the end of §3 the authors commend the designs of Hoke (1974) and Roquemore (1976) when N is hardly larger than p. The numerical comparisons of Atkinson and Tobias (2008) indicate that great care is needed in the choice of these small response surfaces designs. Some that have been suggested in the literature can be surprisingly inefficient, both for cubical and spherical regions.

Table 1 lists the D- and G-efficiency, for a cubical region, of the best of some small designs for $m = 4$ taken from the catalogue of Borkowski (`http://www.math.montana.edu/~jobo/cr/designs.txt`). See also Borkowski and Valeroso (2001). In addition the table includes D-optimum designs found by searching over the 81 points of the 3^4 factorial.

For $N = 15$ the two Hoke designs have identical properties, but have relatively 10% lower D- and G-efficiency than the exact D-optimum design. These three designs share with the others in the table the feature that the D-efficiencies are much higher than the G-efficiencies. For $N = 16$, on the other hand, the D-optimum

Design	N	D-efficiency	G-efficiency
hoke4a1	15	0.8024	0.3913
hoke4a2	15	0.8024	0.3913
D-opt	15	0.8747	0.4522
bd4	16	0.3114	0.0011
Roq416b	16	0.2033	0.0606
D-opt	16	0.8939	0.4167
hoke4a5	19	0.8696	0.3145
D-opt	19	0.9449	0.6900
CCD	25	0.9113	0.7798
D-opt	25	0.9773	0.6891

Table 1: D- and G-efficiencies of some small response surface designs, $m = 4$, cubical region

design far outperforms the Roquemore design and the Box-Draper design that has a G-efficiency of 0.01%, caused by many of the design points being shrunk away from the edges of the design region. The Hoke design for $N = 19$ has a D-efficiency that is again relatively 10% lower than that of the exact D-optimum design, while the ratio of G-efficiencies is much lower at less than one half. Only for $N = 25$ is there a design that challenges the D-optimum design: the CCD with one centre point has, of course, a lower D-efficiency than the D-optimum design, but it does have a higher G-efficiency. This seems to be the only design in these comparisons where the authors' graphical methods would be needed in helping to decide about a design.

Designs are available for some other values of N, for example those of Hoke, which have uniformly poor G-efficiencies. In general, except for a few values of N, there seems no sensible alternative to numerically constructed D- (or G- or I-) optimum designs.

3.3 Costs

Towards the end of §3 the authors discuss the inclusion of costs in experimental design. Fedorov and Leonov (2005) give examples of optimum design incorporating costs and discuss some theory. If the cost of experimenting at x_i is $c(x_i)$ and the total cost of the experiment is constrained to be not greater than C, continuous optimum designs are found as in §2.1 but now with weights $w_i = n_i c(x_i)/C$. See also Atkinson et al. (2007, p. 149). This alternative normalization of designs, originally suggested by Elfving (1952), provides an equivalence theorem and so

should allow modification of the authors' plots to include costs.

3.4 Split-Plot Designs

After many years of comparative neglect, there is now a rapidly expanding literature on the design of split-plot experiments. A book-length treatment of blocked and split-plot designs is Goos (2002). The optimum designs of Goos and Vandebroek (2001) and Goos and Vandebroek (2004) allow for the different costs of changing the levels of the two kinds of factor present in these experiments.

3.5 Compound Designs

In §4.1 the authors consider problems of model robustness that arise because the model is uncertain.

For the jth of h models let the D-efficiency be

$$D_j^{\text{eff}} = \left\{ \frac{|M_j(\xi)|}{|M_j(\xi_j^*)|} \right\}^{1/p_j}. \tag{5}$$

A compound D-optimum design reflecting the weights κ_j of interest in each model then maximizes

$$\Phi(\xi) = \prod_{j=1}^{h} \{D_j^{\text{eff}}\}^{\kappa_j} = \left\{ \frac{|M_j(\xi)|}{|M_j(\xi_j^*)|} \right\}^{\kappa_j/p_j}.$$

On taking logs we see that this is the same as maximizing

$$\sum_{j=1}^{h} \frac{\kappa_j}{p_j} \log |M_j(\xi)| \tag{6}$$

and that

$$d(x,\xi) = \sum_{j=1}^{h} \frac{\kappa_j}{p_j} d_j(x,\xi). \tag{7}$$

An equivalence theorem then applies to this compound design criterion. If ξ^* is the design maximizing (6)

$$\bar{d}(\xi^*) = \sum_{j=1}^{h} \kappa_j.$$

The standard algorithms for the construction of optimum designs can then be used, which may be faster than the genetic algorithm of Heredia-Langner, Montgomery,

Carlyle and Borror (2004). The efficiency of the exact compound designs can be calculated, as can the individual efficiencies (5) and (7) used to provide the authors' plots. Plots for each component $d_j(x, \xi)$ are also a possibility, which would be an extension of the authors' Figure 3. An advantage of this form of compound design over the additive form of Biedermann, Dette, and Zhu (2005) is that calculation of ξ^* does not require knowledge of the individual ξ_j^*. However these designs are, of course, required for the calculation of D-, but not G-, efficiencies.

In practice designs may need to be calculated for several κ_j to find satisfactory component efficiencies. For small h plots of the component efficiencies as a function of the κ_j provide a convenient method of choosing a design. Examples are in Chapter 21 of Atkinson et al. (2007) together with extensions to other forms of compound design, such as DT-optimality, in which interest is in both model discrimination and parameter estimation.

3.6 Model Checking

A disadvantage to the use of compound D-optimum designs is that the design has to have sufficient points of support to fit all models, even the largest. In §4.1 the authors mention potential third-order models. If this were a faint possibility, many design points would have small weights w_i and any exact design would require an excessive value of N. An alternative is the Bayesian model checking design introduced by DuMouchel and Jones (1994) in which prior information about the terms needed for model checking reduces the number of points of support. In the case of checking a second-order model for third-order terms, as this prior information decreases the design would move smoothly from that for the second-order model with one or a few extra design points, to the D-optimum design for the full third-order model. Examples and an equivalence theorem are in Chapter 20 of Atkinson et al. (2007).

3.7 Generalized Linear Models

In the generalized linear model let $\mathrm{E}(y_i) = \mu_i$, with link function $g(\mu_i) = \eta_i$ and linear predictor $\eta_i = \beta^{\mathrm{T}} f(x_i)$. The error distribution is a member of the one-parameter exponential family with $\mathrm{var}\,(y_i) = \phi V(\mu_i)$. Maximum likelihood estimation of β reduces to iteratively re-weighted least squares with, in the notation of the authors' §4.4,

$$v_i = V^{-1}(\mu_i) \left(\frac{d\mu_i}{d\eta_i} \right)^2 .$$

The information matrix for a design ξ, analogous to (2) is thus

$$M(\xi) = F^{\mathrm{T}} W V F, \tag{8}$$

where W and V are both diagonal matrices, the latter with elements v_i. An equivalence theorem then applies to the variance

$$d(x, \xi) = v f^{\mathrm{T}}(x) M^{-1}(\xi) f(x),$$

which extends the authors' definition of the SPV.

The structure of optimum response surface designs is not as clear as that for linear models, since the designs are only locally optimum, depending on the values of the parameters β. However, if the effects β are sufficiently small, the means μ_i will not vary greatly and so the weights v_i will be sensibly constant. Then the information matrix (8) reduces to that for the linear model. Designs for the linear model will therefore be efficient for zero or small effects.

This argument is due to Cox (1988). It seems that the results of Zahran, Myers and Anderson-Cook (2007) mentioned in §4.4 quantify Cox's result for small effects in first-order models. For second-order models, Figure 22.6 of Atkinson et al. (2007) shows how the D-optimum regression design on the points of the 3^2 factorial is gradually distorted as the parameters of a binomial model move away from zero. Their Figure 22.9 shows that for, a gamma model with a Box-Cox link with parameter $\lambda = 0.5$, the distortion of the 3^2 factorial is much less as the parameters change than it is for the binomial model. For the log link the regression design is optimum whatever the parameters of the linear model.

4 Some Symmetrical Optimum Response Surface Designs

An advantage of the methods associated with optimum designs is that they lead to algorithms for the construction of new designs. This section presents a few interesting D-optimum exact designs tabulated by Atkinson and Tobias (2008) that also have good G-efficiency. It might be informative to use fraction of design space plots to compare these designs with others, such as the corresponding G-optimum designs.

The results of Farrell et al. (1968) for cubical regions show that the continuous D-optimum designs are supported on those points of the 3^m factorial with 0, $m-1$ and m non-zero co-ordinates. For the response surface model (1) in two factors these are the points of the 3^2 factorial and $p = 6$. When the design region is continuous and bounded by the square with vertices ± 1, Figure 12.1 of Atkinson et al. (2007) shows that D-optimum designs for $N = 6, 7$ and 8 contain trials not at the points of the 3^2 factorial. However, for $N = 9, 13$ and 14, Atkinson and Tobias (2008) find that the exact D-optimum designs are supported at the factorial points. Figure 1 shows the designs.

1	1	1	2	1	2	2	1	2
1	1	1	1	1	1	1	2	1
1	1	1	2	1	2	2	1	2

Figure 1: Square region, $m = 2$. Designs with good properties for $N = 9, 13$ and 14: $N = 9$, the 3^2 factorial; $N = 13$, a combination of the 3^2 and 2^2 factorials and, $N = 14$, the design for $N = 13$ with an extra centre point

N	D-efficiency	G-efficiency
9	0.9740	0.8276
13	0.9977	0.8718
14	0.9944	0.9606

Table 2: D- and G-efficiencies of D-optimum exact designs of Figure 1 for the second-order model in two factors

All three designs are symmetrical in x_1 and x_2. For $N = 9$ we have the 3^2 factorial, for $N = 13$ the design consists of the 3^2 factorial with replicated corner points, that may also be thought of as the combination of a 2^2 and a 3^2 factorial; the design for $N = 14$ adds a second centre point to the design for $N = 13$. The D-optimum designs for neighbouring values of N lack such symmetry. An advantage of the designs for $N = 13$ and 14 is that the four or five replicated points provide model-free estimates of error variance on 4 or 5 degrees of freedom.

For $m = 3$, there are now $p = 10$ parameters. Atkinson and Tobias (2008) give the properties of D-optimum exact designs for N from 10 to 30. For all $N \neq 10$ the optimum designs when searching over the 5^3 grid are supported at the points of the 3^3 factorial.

The designs for $N = 14, 21$ and 30 are displayed in Figure 2. The designs are shown as 3^2 factorials at each of the levels of a factor arbitrarily labelled x_1. What is immediately noticeable in these designs is the strong symmetric structure. In all three cases the design for x_2 and x_3 is repeated at the high and low levels of x_1. Further, at all levels of x_1 the designs in x_2 and x_3 have a highly symmetric structure, for example a 2^2 factorial with a centre point for $N = 14$. Apart from their good properties an advantage of such designs is that they are easy to specify and so to have performed correctly by unskilled personnel.

Table 3 gives the D- and G-efficiencies of the designs of Figure 2. The D-efficiencies increase steadily from 0.9759 to 0.9995. However, the G-efficiencies

N = 14

x1 = -1			x1 = 0			x1 = 1		
1	0	1	0	1	0	1	0	1
0	1	0	1	0	1	0	1	0
1	0	1	0	1	0	1	0	1

N = 21

x1 = -1			x1 = 0			x1 = 1		
1	1	1	1	0	1	1	1	1
1	0	1	0	1	0	1	0	1
1	1	1	1	0	1	1	1	1

N = 30

x1 = -1			x1 = 0			x1 = 1		
2	1	2	1	0	1	2	1	2
1	0	1	0	2	0	1	0	1
2	1	2	1	0	1	2	1	2

Figure 2: Square region, $m = 3$. Designs with good properties for $N = 13, 21$ and 30 supported on the points of the 3^3 factorial. For each N the figure provides the number of points for each 3^2 factorial in x_2 and x_3 at the three levels of x_1. The labelling of the factors is arbitrary

N	D-efficiency	G-efficiency
14	0.9759	0.8929
21	0.9833	0.8571
30	0.9995	0.9697

Table 3: D- and G-efficiencies of D-optimum exact designs of Figure 2 for the second-order model in three factors

of the designs are much lower than these values; as low as 0.8571 when $N = 21$.

Atkinson and Tobias (2008) also give the properties of D-optimum designs for $m = 4$ for N up to 40. Because of the sparseness of this number of points on the 81 points of the 3^4 factorial, none of the designs has an obvious structure. A challenge is to try to find designs that have good D-efficiency whilst also having good properties as analysed by the authors' plots.

A final comment is that in §§2.1 and 3 the authors mention D- G- and I- (or V)-optimum designs and their construction. Are there cases in which the authors' graphical procedures lead to a different choice of design from that resulting from consideration of these three efficiencies of a design?

References

Atkinson, A. C., A. N. Donev, and R. D. Tobias (2007). *Optimum Experimental Designs, with SAS*. Oxford: Oxford University Press.

Atkinson, A. C. and R. D. Tobias (2008). Optimal experimental design in chromatography. (Submitted).

Biedermann, S., H. Dette, and W. Zhu (2005). Compound optimal designs for percentile estimation in dose-response models with restricted design intervals. In S. Ermakov, V. Melas, and A. Pepelyshev (Eds.), *Proceedings of the 5th St Petersburg Workshop on Simulation*, pp. 143–148. St Petersburg: NII Chemistry University Publishers.

Borkowski, J. J. and E. S. Valeroso (2001). Comparison of design optimality criteria of reduced models for response surface designs in the hypercube. *Technometrics 43*, 468–477.

Box, G. E. P. and D. W. Behnken (1960). Some new 3 level designs for the study of quantitative variables. *Technometrics 2*, 455–475.

Brown, L. D., I. Olkin, J. Sacks, and H. P. Wynn (Eds.) (1985). *Jack Carl Kiefer Collected Papers III*, New York. Wiley.

Cox, D. R. (1988). A note on design when response has an exponential family distribution. *Biometrika 75*, 161–164.

DuMouchel, W. and B. Jones (1994). A simple Bayesian modification of D–optimal designs to reduce dependence on an assumed model. *Technometrics 36*, 37–47.

Elfving, G. (1952). Optimum allocation in linear regression theory. *Annals of Mathematical Statistics 23*, 255–262.

Farrell, R. H., J. Kiefer, and A. Walbran (1968). Optimum multivariate designs. In *Proc. 5th Berkeley Symposium*, Volume 1, Berkeley, CA, pp. 113–138. University of California Press.

Fedorov, V. V. and S. L. Leonov (2005). Response-driven designs in drug development. In M. Berger and W.-K. Wong (Eds.), *Applied Optimal Designs*, Chapter 5, pp. 103–136. New York: Wiley.

Goos, P. (2002). *The Optimal Design of Blocked and Split-plot Experiments*. New York: Springer Verlag.

Goos, P. and M. Vandebroek (2001). Optimal split-plot designs. *Journal of Quality Technology 33*, 436–450.

Goos, P. and M. Vandebroek (2004). Outperforming completely randomized designs. *Journal of Quality Technology 36*, 12–26.