Robustness of response surface designs to missing data

Hleil Alrweili | Stelios Georgiou | Stella Stylianou

School of Science, RMIT University, Melbourne, Victoria, Australia

Correspondence
Stelios Georgiou, School of Science, RMIT University, Melbourne, Victoria, Australia.
Email: stelios.georgiou@rmit.edu.au

Abstract
In this paper, minimax loss response surface designs are constructed. These designs are more robust to one missing design point than the original designs. The proposed designs are compared with the designs in the literature, and they are better in terms of loss and number of runs. Moreover, the new suggestion for the value of \( \alpha \) generates designs not only with less losses but also with higher D-efficiency.

KEYWORDS
central composite designs, D-efficiency, minimax loss criterion, missing data

1 | INTRODUCTION

Missing observations, even in well-planned experiments, can rarely be avoided. Researchers have discovered ways of handling problems associated with missing observations. These include methods either replace the missing values or to apply robust techniques to their analysis. The term “robust” that was first used by Box\(^1\) is often encountered when studying the effects of missing observations. According to Wendelberger,\(^2\) robustness is crucial in the real world of experimentation; thus, the need for experimental designs that are robust in the presence of missing observations arise.

Many researchers have discussed the robustness of statistical designs against missing observations. Hedayat and John\(^3\) and Ghosh\(^4\) considered robust balanced incomplete block (BIB) designs. Herzberg and Andrews\(^5,6\) studied the problem of missing observations in a design from a different angle. They considered the probability of losing an observation at a design point. In a later study, Andrews and Herzberg\(^7\) suggested maximizing the expected value of the determinant of the information matrix to compensate for possible missing observations. Ahmad et al\(^8\) considered multilevel augmented second-order response surface designs and their robustness to missing observations.

In investigating the information contained in an observation, Ghosh\(^9\) stated that in most designs, some design points were more informative than others and that the effects of one or more design points, when missed, attracted a higher loss in efficiency. Other economical designs requiring fewer design points were constructed and studied for their robustness to missing observations by Ahmad and Akhtar.\(^10\) Angelopoulos et al\(^11\) constructed new central composite designs (CCDs) that are more economical than the classical CCDs that were introduced by Box and Wilson\(^12\) since the newer designs were built with fewer runs. Using their designs, they found that the minimum number of runs for a four-factor design was 18. Another economical design was constructed by Georgiou et al.\(^13\) They constructed a class of composite designs with a higher D-value (see the study of Lucas\(^14\) ) than Angelopoulos's designs but with the same number of runs.

Various robust-to-missing-observations criteria have been proposed in the literature. Akhtar and Prescott\(^15\) developed the minimax loss criterion and applied it to the evaluation and generation of CCDs. They found that, when they applied minimax loss criterion on the classical CCDs with four factors, the number of runs was 26. Since then, this criterion has been adopted by many researchers including Ahmad and Gilmour,\(^16\) who used it to study the robustness of subset designs. In addition, the hat matrix, also known as the projection matrix, performs a major role in quantifying the effect of removing one or more observations from a design.\(^17\) Box and Draper\(^18\) pointed out the connection between the diagonal
elements of a hat matrix and a design's robustness and applied their related criterion to CCDs. Akram\textsuperscript{19} utilized the hat matrix (through its compounds) in studying the robustness of CCDs up to three missing observations.

Each robust-to-missing-observations criterion has its limitation(s). One limitation is that they cannot be applied with the remaining runs if these runs are not enough to estimate the parameters of interest. Another limitation is that each criterion focuses on one area of interest such as loss, D-efficiency, estimation capacity, and so on. Therefore, more flexible robust-to-missing-observations criteria may be required to circumvent such restrictions. In the literature, there is nothing on the effect of applying minimax loss criterion on the latest CCDs that are constructed with a minimum number of runs. This paper integrates the minimax loss criterion with the latest CCDs proposed by Georgiou et al\textsuperscript{13} denoted as GSA, and Angelopoulos et al\textsuperscript{11} denoted as AEK, to achieve more robust designs in the case of missing observations. The GSA and AEK designs are used in this paper because these designs have less runs than the classical CCDs that were evaluated by Akhtar and Prescott.\textsuperscript{15}

In section 2, the construction of the GSA and AEK designs are discussed. In section 3, the minimax loss criterion is discussed. Section 4 details the methodology. The new results and a comparison of our findings are presented in section 5.

2 | GSA AND AEK

In this section, we present the proposed construction methods of the GSA and AEK designs that are used in this paper.

2.1 | GSA designs

The GSA designs are constructed as a three-part design matrix $X$ that would be suitable for fitting the second-order response surface as follows:

1. For the factorial part, select a suitable number of balanced columns from a known two-level design with good properties. For example, one may use columns from a Hadamard matrix of order $n_f$, from a fractional factorial design with $n_f$ runs, from a supersaturated design with a run size of $n_f$, or from other suitable two-level design where $n_f$ is to be selected. They called this part of the design $F$.
2. Instead of using the traditional axial points, they suggested using vertices of a hypercube that circumscribe the sphere of zero center and radius $a$. This approach includes the axial points as a special case. They called this part of the design $V$.
3. Select a number of center points $n_c$, and this part of the design is denoted as $C$.

The desirable GSA design matrix is as follows:

$$X = (F', V', C', -V)$$

Example 1. Suppose we wish to construct a six-factor GSA design. Select six columns for part $F$ from the Hadamard matrix of order 20, and for part $V$, select six columns of the weighing matrix of order 8 ($W_8$) multiplied by $a$. More details on weighing matrices can be found in Koukouvinios and Seberry.\textsuperscript{20}

$$W_8 = \begin{bmatrix}
0 & -1 & -1 & -1 & -1 & 0 & 0 \\
1 & 0 & -1 & 1 & -1 & 1 & 0 \\
1 & 1 & 0 & 1 & 0 & 1 & 1 \\
1 & -1 & 1 & 0 & 0 & 1 & 1 \\
1 & 1 & 0 & 0 & -1 & 1 & -1 \\
1 & -1 & 0 & 1 & 0 & -1 & -1 \\
0 & 0 & 1 & -1 & -1 & 1 & 0 \\
0 & 0 & 1 & -1 & 1 & 1 & 0
\end{bmatrix}.$$

2.2 | AEK designs

The AEK designs are constructed by an exhaustive computer search. The aim was to find the minimum number of suitably selected points that are needed for the fractional factorial part (four factors to nine factors) so that the designs estimate all the model parameters. The classical method of CCDs is used for the axial part. Table 1 present the general structure of the AEK designs.

Example 2. Using a computer search for the factorial part, the authors provide the following design of six factors that maximize the criterion.
TABLE 1  General structure of GSA designs with $k = 6$ factors

\[
\begin{pmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 0 & -\alpha & -\alpha & -\alpha & -\alpha & 0 & 0 & \alpha & \alpha & \alpha & \alpha & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 0 & -\alpha & -\alpha & -\alpha & 0 & -\alpha & 0 & \alpha & -\alpha & \alpha & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 0 & -\alpha & -\alpha & -\alpha & 0 & 0 & -\alpha & -\alpha & -\alpha & 0 & 0 & \alpha \\
1 & 1 & 1 & 1 & 1 & 1 & 0 & -\alpha & -\alpha & -\alpha & 0 & 0 & 0 & -\alpha & -\alpha & -\alpha & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 0 & -\alpha & -\alpha & -\alpha & 0 & 0 & 0 & 0 & -\alpha & -\alpha & -\alpha & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 0 & -\alpha & -\alpha & -\alpha & 0 & 0 & 0 & 0 & 0 & -\alpha & -\alpha \\
1 & 1 & 1 & 1 & 1 & 1 & 0 & -\alpha & -\alpha & -\alpha & 0 & 0 & 0 & 0 & 0 & 0 & -\alpha & -\alpha \\
1 & 1 & 1 & 1 & 1 & 1 & 0 & -\alpha & -\alpha & -\alpha & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\alpha & -\alpha \\
\end{pmatrix}
\]

3  MINIMAX LOSS CRITERION

Akhtar and Prescott\textsuperscript{15} studied the reduction in the determinant of the information matrix because of missing observations. As previously mentioned, they developed the minimax loss criterion to find designs robust to missing observations. The effect of missing observations is quantified using the criterion of the loss function proposed by Akhtar and Prescott.\textsuperscript{15}

More specifically, the loss function gives a measure of the relative reduction in the determinant of the information matrix associated with the complete design in the presence of missing observations. More studies on the loss function consider second-order CCDs differing in the number of control variables, configurations of the factorial, and axial and centre portions.

For a $p$ parameters, $p = ((K + 1)(K + 2))/2$, second-order model

\[
y = \beta_0 + \sum_{i=1}^{K} \beta_i x_i + \sum_{i<j} \beta_{ij} x_i x_j + \sum_{i=1}^{K} \beta_i x_i^2 + \epsilon,
\]

written in matrix form as follows:

\[
y = X\beta + \epsilon.
\]

where $K$ is the number of factors, $y$ denotes the vector of observations, $X$ denotes the model matrix, $\beta$ denotes the vector of unknown parameters that are estimated based on $N$ uncorrelated observations and where $\epsilon$ denotes the random error component assumed to be normally and independently distributed with a mean of zero and a constant variance. The $N \times K$ design matrix, $D$, comprising $N$ design points, is expanded into $N \times p$ model matrix $X$. The columns of $X$ are obtained using the design matrix and the full second-order model, where the number of model parameters $p$ equals $((K + 1)(K + 2))/2$.

When $m$ observations in the design are missing, the model matrix reduces by $m$ rows. As expected, the information matrix, $X'X$, for the reduced design will be different from the information matrix, $X'X$, for the complete design. If $X_m$ is the $m \times p$ matrix of the $m$ missing rows corresponding to the $m$ missing observations, then the information matrix may be expressed as follows:

\[
X'X = X'_mX_m + X'_rX_r.
\]

By post multiplication,

\[
(X'X)(X'X)^{-1} = [X'_mX_m + X'_rX_r](X'X)^{-1}.
\]

This implies

\[
I_p = (X'_mX_m)(X'X)^{-1} + (X'_rX_r)(X'X)^{-1};
\]

$I_p$ is an identity matrix of order $p$. By rearrangement, we have

\[
(X'_rX_r)(X'X)^{-1} = I_p - (X'_mX_m)(X'X)^{-1}.
\]

Post multiplying by $(X'X)$ yields

\[
(X'_rX_r)(X'X)^{-1}(X'X) = [I_p - (X'_mX_m)(X'X)^{-1}](X'X)
\]

\[
\Rightarrow (X'_rX_r) = [I_p - (X'_mX_m)(X'X)^{-1}](X'X),
\]

so the determinant will be

\[
|X'_rX_r| = \left| \left| I_p - (X'_mX_m)(X'X)^{-1} \right| \right| |(X'X)|.
\]

This implies

\[
\frac{|X'_rX_r|}{|X'X|} = \left| \left| I_p - (X'_mX_m)(X'X)^{-1} \right| \right|.
\]
Akhtar and Prescott\textsuperscript{15} call $[I_p - (X_m'X_m)(X'X)^{-1}]$ the diagonal element of the $m^{th}$ compound of $(I - R)$ where $R$ is the hat matrix according to the $m$ missing observations and $I$ is an identity matrix of the same dimension as $R$.

For a square matrix $H$ of order $n$, the $m^{th}$ compound of $H$ is a square matrix of order $\binom{n}{m}$ the elements of which are the minors of determinant $|H|$ of order $m$. For a particular number of design points $n$ and number of parameters $p$ to be estimated, the sum of diagonal elements of the $m^{th}$ compound is constant. For more detail, see Aitken and Rutherford.\textsuperscript{21}

Andrews and Herzberg\textsuperscript{7} presented a relation between $|X'X|$ and $|X'X|$ by using identities on the expansion of a bordered determinant (see the study of Rao\textsuperscript{22}) as follows:

$$\|(X'X_o)\| = \|(X'X)\| A\ldots,$$ \hspace{1cm} (10)

where $A\ldots$ is the corresponding diagonal element of the $m^{th}$ compound of $(I - R)$ matrix and $R$ is the symmetric matrix $X'X^{-1}X'$ of order $N$. Specifically, $A_u$ is the $u^{th}$ diagonal element of the first compound of $(I - R)$ matrix, $A_{uv}\ldots$ is the $uv^{th}$ diagonal element of the second compound of $(I - R)$ matrix and so on.

For $m$ missing observations, the determinant of the complete information matrix may be obtained as follows:

$$\|(X'X)\| = \|(X_m'X_m)\| + \|(X'X_o)\|.$$ \hspace{1cm} (11)

The relative loss because of any set of $m$ missing observations is defined as follows:

$$l_{uv} = \frac{|(X'X) - \|(X'X_o)\|}{\|(X'X)\|}; \hspace{1cm} 0 \leq l_{uv} \leq 1.$$ \hspace{1cm} (12)

It was found that

$$l_{uv} = 1 - \frac{|(X'X_o)|}{\|(X'X)\|} = 1 - A\ldots = 1 - \left|I - (X_m'X_m)(X'X)^{-1}\right| \hspace{1cm} \text{(See Akram)}\textsuperscript{19}. \hspace{1cm} (13)$$

Under D-optimality, the determinant $|\|(X'X)\|$ is maximized. It is well established for CCDs that the determinant $|\|(X'X)\|$ is an increasing function of the axial distance $\alpha$ and is maximized at $\alpha = \infty$. The value of $\alpha$ specifies the location of the axial point and may be chosen to satisfy various design conditions such as rotatability and orthogonal blocking.

In the presence of missing observations, the determinant of the information matrix $|\|(X'X)\|$ of a complete design is reduced to $\|(X'X_o)\|$. It is essential that this reduction is as small as possible. Thus, the relative loss, $l_{uv}\ldots$ needs to be minimized, and this is possible if

$$A\ldots = \left|1 - \left[I - (X_m'X_m)(X'X)^{-1}\right]\right| = \frac{|(X'X_o)|}{\|(X'X)\|} \hspace{1cm} (14)$$

tends to one. The most practicable way to minimize the losses is to minimize the maximum loss because of a set of $m$ missing observations. Minimizing the maximum loss is the idea behind the minimax loss criterion of Akhtar and Prescott.\textsuperscript{15}

### 4 | METHODOLOGY

#### 4.1 | The choice of $\alpha$ for the new designs

Alpha ($\alpha$) is used to satisfy or maximize some conditions such as the rotatability of the design or to maximize the determinant of the information matrix. Fang and Mukerjee\textsuperscript{23} discussed the consequences of rescaling $\alpha$ by restricting themselves to a cuboidal region \([-1, 1]\). They concluded that $\alpha = 1$ is the best choice under the D-criterion. Here, the condition of the cuboidal design region was relaxed, and the designs were constructed by finding a new value of $\alpha$ using the minimax loss criterion presented by Akhtar and Prescott.\textsuperscript{15} This criterion is propitious for the construction of designs that are more robust to missing observations. The value of $\alpha$ was chosen so that the maximum loss of a missing design point is minimized.
4.2 The detailed procedure for constructing the proposed designs

We now present the detailed procedure for constructing the proposed designs.

Algorithm

Step 1. Choose an AEK or GSA design \( X = \begin{bmatrix} F & V \\ -V & -V \end{bmatrix} \), where \( F \) is the factorial part and \( V \) is the axial part. Note that center points may be added to the chosen design.

Step 2. Calculate the loss function \( L_f, L_a, \) and \( L_0 \) based on \( X \) for the second order model (1), where \( L_f \) is the loss for a missing factorial point, \( L_a \) is the loss for a missing axial point, and \( L_0 \) is the loss for a missing centre point.

Step 3. Find \( \alpha \) that minimizes the maximum loss of \( \{ L_f, L_a, L_0 \} \).

Step 4. Substitute the new \( \alpha \) in the chosen design.

5 THE NEW RESULTS

In this section, we describe the new results we obtain by exhaustively searching the values of \( \alpha \) that minimize the maximum loss.

Example 3. An example of how to find \( \alpha \) that minimizes the maximum loss for 6 factors with one missing observation is given in Table 2. The general structure of the design matrix is presented in Table 3.

It can be seen in Figure 1 that \( L_f \) is decreasing and \( L_a \) is increasing due to the function of \( \alpha \) with a negative and positive slope on the graph. In order to compute the value of \( \alpha \) at which the maximum loss is minimized, function \( L_f \) is equated with \( L_a \), and it was observed that 1.335 was the minimum value of \( \alpha \) at which \( L_f \) is equal to one of the other functions. Thus, the maximum loss of missing one design point will be minimized at \( \alpha = 1.335 \) with loss = 0.89.

| \( \alpha \) | \( L_f \) | \( L_a \) | \( L_0 \) | \( \alpha \) | \( L_f \) | \( L_a \) | \( L_0 \) |
|---|---|---|---|---|---|---|---|
| 0.5 | 0.9769 | 0.8530 | 0.0659 | 1.4 | 0.8941 | 0.9042 | 0.2278 |
| 0.55 | 0.9729 | 0.8633 | 0.0702 | 1.45 | 0.8920 | 0.9092 | 0.2136 |
| 0.6 | 0.9689 | 0.8712 | 0.0755 | 1.5 | 0.8904 | 0.9139 | 0.1984 |
| 0.65 | 0.9647 | 0.8772 | 0.0818 | 1.55 | 0.8892 | 0.9184 | 0.1834 |
| 0.7 | 0.9605 | 0.8814 | 0.0895 | 1.6 | 0.8881 | 0.9226 | 0.1693 |
| 0.75 | 0.9562 | 0.8843 | 0.0986 | 1.65 | 0.8873 | 0.9264 | 0.1565 |
| 0.8 | 0.9518 | 0.8860 | 0.1096 | 1.7 | 0.8866 | 0.9298 | 0.1450 |
| 0.85 | 0.9473 | 0.8884 | 0.1226 | 1.75 | 0.8859 | 0.9329 | 0.1349 |
| 0.9 | 0.9425 | 0.8872 | 0.1378 | 1.8 | 0.8853 | 0.9358 | 0.1260 |
| 0.95 | 0.9376 | 0.8870 | 0.1551 | 1.85 | 0.8847 | 0.9384 | 0.1181 |
| 1 | 0.9324 | 0.8865 | 0.1743 | 1.9 | 0.8841 | 0.9409 | 0.1113 |
| 1.05 | 0.9268 | 0.8861 | 0.1943 | 1.95 | 0.8835 | 0.9431 | 0.1052 |
| 1.1 | 0.9211 | 0.8861 | 0.2138 | 2 | 0.8829 | 0.9452 | 0.0999 |
| 1.15 | 0.9154 | 0.8867 | 0.2306 | 2.05 | 0.8823 | 0.9472 | 0.0952 |
| 1.2 | 0.9098 | 0.8884 | 0.2426 | 2.1 | 0.8817 | 0.9490 | 0.0911 |
| 1.25 | 0.9048 | 0.8911 | 0.2482 | 2.15 | 0.8811 | 0.9507 | 0.0874 |
| 1.3 | 0.9004 | 0.8949 | 0.2469 | 2.2 | 0.8805 | 0.9524 | 0.0841 |
| 1.335 | 0.8978 | 0.8978 | 0.2423 | 2.25 | 0.8799 | 0.9539 | 0.0812 |
| 1.35 | 0.8969 | 0.8993 | 0.2396 |
FIGURE 1 Loss for different values of $\alpha$ for the GSA design with $K = 6, n_c = 4$ [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 4 Comparison of maximum loss and number of runs (N) of different designs with $K$ factors and the suggested $\alpha$ values

| $K$ | $p$ | CCD   | SCD   | AP    | AEKM  | GSAM   |
|-----|-----|-------|-------|-------|-------|--------|
| 4   | 15  | .5684(29) | .9198(21) | .7003(41) | .9225(21)* | .8570(21)* |
|     |     | $\alpha = 1.9$ | $\alpha = 1.4$ | $\alpha = 1.03$ | $\alpha = 1.57$ | $\alpha = 2.99$ |
| 5   | 21  | .8749(31) | .9801(27) | .8368(41) | .9581(27)* | .9649(27)* |
|     |     | $\alpha = 3.4$ | $\alpha = 2.6$ | $\alpha = 1.09$ | $\alpha = 3.38$ | $\alpha = 3.77$ |
| 6   | 28  | .6231(49) | .9866(33) | .9203(41) | .9592(35)* | .8978(41)* |
|     |     | $\alpha = 2.7$ | $\alpha = 2.1$ | $\alpha = 1.06$ | $\alpha = 3.8$ | $\alpha = 3.35$ |
| 7   | 36  | .4237(83) | .9648(43) | .9894(41) | .9858(41)* | .9910(43)* |
|     |     | $\alpha = 2.1$ | $\alpha = 3.1$ | $\alpha = 1.06$ | $\alpha = 3.45$ | $\alpha = 3.57$ |

$^a$States $n_c = 1$. $^b$States $n_c = 3$. $^c$States $n_c = 5$. Central composite designs introduced by Box and Wilson,12 and their $\alpha$ values were investigated by Akhtar and Prescott.15 Small composite designs introduced by Draper and Lin,24 and their $\alpha$ values are studied here for the first time. Augmented pairs designs introduced by Morris,25 and their $\alpha$ values were investigated by Ahmad et al.8 These designs introduced by Angelopoulos et al,11 and their $\alpha$ values are studied here for the first time. These designs introduced by Georgiou et al,13 and their $\alpha$ values are studied here for the first time.

Table 4 shows the values of maximum loss for one missing design point with the number of runs in parentheses and the values of $\alpha$ that we used to construct new designs that are robust against a missing observation. From Table 4, it is obvious that the newly proposed designs, AEKM, which are the minimax designs derived from AEK and GSAM that are the minimax designs derived from GSA, perform better in comparison to at least one of the three classical designs that are used in the literature on CCD, small composite design (SCD), and augmented pair (AP) designs. For example, the five-factor design of GSA with 26 runs and one centre point has the lowest value of maximum loss for one missing point of all the designs in Table 4. Furthermore, it had a fewer number of runs, which makes it an economical and robust design for one missing observation. Although the CCD is superior in most cases, the newly proposed designs show improvement over one or more of the other two designs. The designs given in bold face provide the minimum number of runs for estimating $p$ parameters with less loss than the designs in the literature in the case of one missing observation. In summary, the new
TABLE 5 | Relative D-efficiencies of the robust designs (using the new selection of $\alpha$ [$\alpha_{\text{new}}$]) in relation to the original designs (having the suggested $\alpha$ in the literature [$\alpha_{\text{lit}}$])

| $n_C$ | $K=4$ | $K=5$ | $K=6$ | $K=7$ |
|-------|-------|-------|-------|-------|
|       | GSAM  | AEKM  | GSAM  | GSAM  |
|       | $N=16$| $N=22$| $N=24$| $N=26$|
|       | $\alpha_{\text{new}}=3.14$| $\alpha_{\text{new}}=1.36$| $\alpha_{\text{new}}=3.52$| $\alpha_{\text{new}}=1.33$|
|       | $\alpha_{\text{lit}}=2.4$ | $\alpha_{\text{lit}}=2.6$ | $\alpha_{\text{lit}}=2.6$ | $\alpha_{\text{lit}}=3$ |
| 1     | 1.14  | 1.16  | 2.7   | 2.19  |
| 5     | 1.20  | 1.19  | 2.84  | 2.29  |
|       | 1.12  | 2.48  | 1.08  | 1.1   |

TABLE 6 | Comparison of the D-efficiency of the proposed designs with the existing designs in the literature

| $K$ | $p$ | CCD | SCD | AP | AEKM | GSAM |
|-----|-----|-----|-----|----|------|------|
| 4   | 15  | N = 24 | N = 16 | N = 36 | N = 16 | N = 16 |
|     |     | a = 1.9 | a = 1.4 | a = 1.03 | a = 1.75 | a = 3.14 |
| 5   | 21  | N = 26 | N = 22 | N = 36 | N = 22 | N = 26 |
|     |     | a = 3.4 | a = 2.6 | a = 1.09 | a = 3.38 | a = 1.36 |
|     |     | N = 26 | N = 36 | N = 36 | N = 28 | N = 36 |
|     |     | a = 2.7 | a = 2.1 | a = 1.06 | a = 3.8 | a = 1.33 |
| 6   | 28  | N = 44 | N = 28 | N = 36 | N = 30 | N = 36 |
|     |     | a = 2.7 | a = 2.1 | a = 0.81 | a = 3.45 | a = 3.57 |
| 7   | 36  | N = 78 | N = 36 | N = 36 | N = 38 | N = 38 |
|     |     | a = 2.1 | a = 3.1 | a = 3.1 | a = 3.45 | a = 3.57 |

results on Table 4 are $K = 4$ with 21 runs, $K = 5$ with 27 runs and 31 runs, and $K = 6$ with 35 and 41 runs. Also, there is a new improved design for $K = 7$ with 41 runs.

Table 5 shows the relative D-efficiency of the robust designs with the new $\alpha$ values that are presented in bold in Table 4. Following the literature, comparison is presented in the cases of one centre point and five centre points. Since we investigated all possible values of $\alpha$ to find the one that gives the less minimax loss when a design point is missing, then it makes sense to see whether we lost D-efficiency by trying to optimize the design based on the loss function. Thus, we think that a comparison of the relative D-efficiency of designs with the $\alpha$ values suggested in the literature and the new $\alpha$ values suggested here is relevant. Table 5 shows that the designs with minimum losses, AEKM and GSAM, also perform better in terms of the relative D-efficiency when they are compared with the classical AEK and GSA designs (with the $\alpha$ used in the literature).

The D-efficiency of a design is defined as follows:

$$D\text{-efficiency} = \left| X'X \right|^{1/p}, \quad (15)$$

where $X$ is the corresponding model matrix that is defined in (2).

The relative D-efficiency of GSAM is defined to be

$$\text{relative D-efficiency} = \left| X'X \right|_{\text{GSAM}} / \left| X'X \right|_{\text{GSA}}^{1/p}. \quad (16)$$

The relative D-efficiency of AEKM is defined as follows:

$$\text{Relative D-efficiency} = \left| X'X \right|_{\text{AEKM}} / \left| X'X \right|_{\text{AEK}}^{1/p}. \quad (17)$$

The $\alpha$ values in Table 6 are selected so that the design gives the minimax loss. For the $\alpha$ values of CCD and AP that give the minimax loss, see Akhtar and Prescott\textsuperscript{15} and Ahmad et al.\textsuperscript{8} respectively. For SCD, AEKM, and GSAM, we investigated exhaustively and found the values of $\alpha$ that give the minimax loss. So Table 6 compares the D-efficiency of different designs using the $\alpha$ values that make them robust against a missing observation. The purpose is that all the designs in Table 6 are robust against a missing observation, so D-efficiency gives a secondary evaluation of these robust designs. It resulted in
most of the generated robust designs, in particular the AEKM designs, having higher D-efficiency than the designs in the literature. For instance, the six-factor design of the AEKM appears to have higher D-efficiency, even though it has less runs. The six-factor design is an improvement over all the known designs in the literature with same number of runs. Moreover, this design overtakes the other known designs in the literature even if those have more runs. In addition, if we compare the five-factor design of AEKM with AP and SCD, we can see that AEKM’s D-efficiency is higher than both AP and SCD even though its number of runs is equal to SCD and less than AP design. GSAM designs have less D-efficiency than the other designs except in the case of five factors and 26 runs where this design is superior to all the designs in the literature.

6 | DISCUSSION

Robust designs are essential in the case of missing observations as even the most well-planned experiments are susceptible to missing data. Choosing a response surface design to fit certain kinds of models is a difficult task. However, a design that performs well when an experimental observation is missing is always the most desirable. In this article, new designs were presented as a variation of the designs provided in GSA and AEK, which are more robust in the case of missing observations. The modified designs used the minimax loss criterion to choose a value of \( \alpha \) for a whole set of axial parts in the experimental region. The newly proposed designs showed better performance than the GSA and AEK designs and occasionally performed better than CCD, SCD, and AP when the restriction to a cuboidal experimental region was relaxed. These designs are also comparable with standard CCDs because of their five levels for each factor. CCDs are still superior in most cases, but because of a significant reduction in the design size, for larger \( K \), the newly proposed designs may be more preferable in some cases. Even though the newly proposed designs are introduced as single-stage designs, their structure suggests their potential for use in sequential experimentation as well.

Additional interested problems in this field would be to investigate the changes in losses and \( \alpha \) value in the case of more than one missing observations. That interesting and related problem might be of particular interest and will be a generalization of this work. Furthermore, rotatability is another additional criterion that can be impeded in studying response surface designs but this will requires extended effort as different missing observations may resulted at the same losses but provide different rotatability properties.

Generally, when researchers are dealing with missing observations, they need to be extremely careful trying to handle simultaneously optimization of several criteria. The multiple stage analysis will be affected by the initial screening stage and the overall efficiency will also be modified.

ORCID

Stelios Georgiou https://orcid.org/0000-0001-5096-0067

REFERENCES

1. Box G. Non-normality and tests on variance. *Biometrika*. 1953;40:318-335.
2. Wendelberger J. Uncertainty in designed experiments. *Qual Eng*. 2010;22(2).
3. Hedayat A, John P. Resistant and susceptible BIB designs. *Ann Stat*. 1974;1:148-158.
4. Ghosh S. Robustness of BIB designs against non availability of data. *J Stat Plan Infer*. 1982a;6:29-32.
5. Herzberg A, Andrews D. The robust design of experiments. *Bull Int Stat Inst*. 1975;45(II):370-374.
6. Herzberg A, Andrews D. Some considerations in the optimal design of experiments in non-optimal situations. *J R Stat Soc B*. 1976;38:284-289.
7. Andrews DF, Herzberg AM. The robustness and optimality of response surface designs. *J Stat Plan Infer*. 1979;3:249-257.
8. Ahmad T, Akhtar M, Gilmour SG. Multilevel augmented pairs second-order response surface designs and their robustness to missing data. *Comput Stat Theory and Methods*. 2012;41(3):437-452.
9. Ghosh S. Information in an observation in robust designs. *Comput Stat Theory Methods*. 1982b;11:1173-1184.
10. Ahmad T, Akhtar M. Efficient response surface designs for the second-order multivariate polynomial model robust to missing observation. *J Stat Theory Pract*. 2015;9(2):361-375.
11. Angelopoulos P, Evangelaras H, Koukouvinos C. Small, balanced, efficient and near-rotatable central composite designs. *J Stat Plan Infer*. 2009;139:2010-2013.
12. Box G, Wilson K. On the experimental attainment of optimum conditions. *J R Stat Soc Ser B*. 1951;13:1-45.
13. Georgiou S, Stylianou S, Aggarwal M. A class of composite designs for response surface methodology. *Comput Stat Data Anal*. 2014;71:1124-1133.
AUTHOR BIOGRAPHIES

Hleil Alrweili is a PhD student in Statistics at the School of Science, RMIT University, Melbourne, Australia. He completed his research, and he is currently at the stage of writing his PhD thesis. His research interests include Applied Statistic, Experimental Designs, and Statistical Quality Control.

Stelios Georgiou is an Associate Professor at the School of Science, RMIT University, Melbourne, Australia. He received his PhD from the National Technical University of Greece in the field of Combinatorics and Statistics. He publishes extensively in numerous international journals and is a member of the American Statistical Society and the Australian Statistical Society. His current research interest include Applied Statistics, Design and Analysis of Experiments, Statistical Quality Control, and Combinatorics.

Stella Stylianou is a Senior Lecturer at the School of Science, RMIT University, Melbourne, Australia. She received her PhD from the National Technical University of Greece in the field of Statistics. She has published in numerous accredited peer-reviewed journals and has presented her research at several National and international conferences. Her research interest include Biostatistics, Applied Statistics, and Design of Experiments.

How to cite this article:  Alrweili H, Georgiou S, Stylianou S. Robustness of response surface designs to missing data. Qual Reliab Engng Int. 2019;35:1288–1296. https://doi.org/10.1002/qre.2524