Quality Engineering
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lqen20

Selecting a Best Two-Level 16-Run Screening Design from the Catalog of Nonisomorphic Regular and Nonregular Designs for Six to Eight Factors

Lu Lu [a] , Mark E. Johnson [b] & Christine M. Anderson-Cook [c]
[a] Department of Mathematics and Statistics , University of South Florida , Tampa , Florida
[b] Department of Statistics , University of Central Florida , Orlando , Florida
[c] Statistical Sciences Group , Los Alamos National Laboratory , Los Alamos , New Mexico
Published online: 27 May 2014.

Quality Engineering, 26:269–284, 2014

ISSN: 0898-2112 print/1532-4222 online
DOI: 10.1080/08982112.2013.854903

Taylor & Francis
Taylor & Francis Group

Selecting a Best Two-Level 16-Run Screening Design from the Catalog of Nonisomorphic Regular and Nonregular Designs for Six to Eight Factors

Lu Lu[1],
Mark E. Johnson[2],
Christine M. Anderson-Cook[3]

[1]Department of Mathematics and Statistics, University of South Florida, Tampa, Florida
[2]Department of Statistics, University of Central Florida, Orlando, Florida
[3]Statistical Sciences Group, Los Alamos National Laboratory, Los Alamos, New Mexico

ABSTRACT When exploring first-order models including two-factor interactions for six to eight factors using a 16-run design, there are many possible model choices. Building on the Johnson and Jones (2010) catalog of the nonisomorphic regular and nonregular design alternatives, we summarize which of these design options are most promising based on two common design criteria. The Pareto fronts based on the criteria $E(s^2)$ and $tr(\boldsymbol{AA'})$ suggest that only a handful of the possible designs should be considered further, and the best design depends on the relative emphasis given each of the two criteria. This article considers each case of six, seven, and eight factors for 16-run two-level designs and provides numerical and graphical comparisons between the alternatives to highlight the merits of the leading candidates.

KEYWORDS alias patterns, design generators, desirability functions, nondominated designs, Pareto front optimization, supersaturated designs

INTRODUCTION

In many design scenarios, tight budgetary constraints restrict the size of a design to be relatively small compared to the number of design factors of interest. For example, the classical fractional factorial resolution IV design with defining relation I = ABCE = BCDF = ADEF is a common choice for a screening experiment with 16 runs and six factors. This design ensures that all main effects are not confounded with other main effects or any two-factor interactions. However, this design does not allow separate estimation of all two-factor interactions, since these are confounded with several other two-factor interactions. Jones and Montgomery (2010) proposed some nonregular designs with no complete confounding between any main effects or two-way interactions. Johnson and Jones (2010) provided a catalog of defining equations for all orthogonal nonisomorphic regular and nonregular designs, which are considered as logical candidates for constructing 16-run designs for six, seven, and eight factors.

Address correspondence to Lu Lu, 4202 E. Fowler Avenue, Tampa, FL 33620. E-mail: icyemma@gmail.com

Using the Johnson and Jones (2010) catalog, this article evaluates the individual design performance of all candidates based on two common general criteria and eliminates the noncontenders for decision making based on using a Pareto front approach (Lu et al. 2011, 2012). The remaining promising choices are compared based on performance given different potential prioritization of the criteria, which should be governed by the goals of the study.

Two metrics for quantifying omnibus design characteristics are considered. To select a screening design with k factors, the main goal is to identify active design factors. We focus on factors influencing the response through main effects and two-factor interactions in a first-order model with all $\binom{k}{2}$ two-way interactions:

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i X_i + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \beta_{ij} X_i X_j + \varepsilon. \qquad [1]$$

Hence, we want not only good estimation of main effects but also the ability to estimate any potentially active two-factor interactions. In our case, the number of two-factor interactions to be explored in model [1] grows as the number of factors increases (six factors have 15 interactions, seven factors have 21, and eight factors have 28 interactions). Because of the limited number of runs, we cannot estimate all terms in model [1]. There will be complete (from regular fractional factorial designs) or partial (from nonregular designs) aliasing, which complicates the estimation procedures. Since we do not know in advance which two-factor interactions will be present in the final model, the general design characteristics look at the potential impact of any combination of these terms being active. In this case, we can think of our screening designs as supersaturated designs (Booth and Cox 1962), which have fewer runs than effects to be estimated in the proposed model.

We assume the matrix form for model [1] is given by

$$\boldsymbol{y} = \beta_0 \mathbf{1} + \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \qquad [2]$$

In [2], $\boldsymbol{y}$ is a vector of $N = 16$ observations for the response, $\boldsymbol{X}$ is an $N \times f$ matrix containing columns for the k main effects and $\binom{k}{2}$ two-factor interactions, $\boldsymbol{\beta}$ is the vector of model parameters excluding the intercept, $\mathbf{1}$ is an $N \times 1$ vector with all entries being 1, β_0 is the intercept, and $\boldsymbol{\varepsilon}$ is the vector for the experimental errors. Booth and Cox (1962) suggested that a good supersaturated design should have off-diagonal elements of $\boldsymbol{X'X}$ as small as possible and introduced the $E(s^2)$ criterion for selecting supersaturated designs that minimizes

$$E(s^2) = \frac{2}{f(f-1)} \sum_{i<j} s_{ij}^2, \qquad [3]$$

where s_{ij} is the element in the ith row and jth column of $\boldsymbol{X'X}$. If off-diagonal elements are small, then this corresponds to lower correlations between estimated model coefficients.

The second criterion is the $tr(\boldsymbol{AA'})$ proposed by Bursztyn and Steinberg (2006) for quantifying potential impact from model misspecification. In the selection of a screening design, we assume that the primary interest lies in good estimation of main effects in the specified model

$$\boldsymbol{y} = \boldsymbol{X_1\beta_1} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{X}_1$ is the design matrix containing only main effects and $\boldsymbol{\beta}_1$ is the vector of corresponding model parameters including the intercept, if applicable. However, it is also thought that some subset of the two-factor interactions is potentially active. Hence, we want some protection for the estimation of the specified model by seeking designs with minimal bias in the estimated parameters, if this model is incorrect and some two-way interactions exist. Let

$$\boldsymbol{y} = \boldsymbol{X_1\beta_1} + \boldsymbol{X_2\beta_2} + \boldsymbol{\varepsilon}$$

denote the larger model with the additional term $\boldsymbol{X}_2\boldsymbol{\beta}_2$ containing all possible two-factor interactions. The bias of the least squares estimate of $\boldsymbol{\beta}_1$ is

$$E\left(\widehat{\boldsymbol{\beta}}_1\right) - \boldsymbol{\beta}_1 = \left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2\boldsymbol{\beta}_2 = \boldsymbol{A\beta}_2,$$

where $\boldsymbol{A} = \left(\boldsymbol{X}_1'\boldsymbol{X}_1\right)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2$ is the alias matrix measuring the degree of bias for estimated model parameters due to the existence of active terms in $\boldsymbol{X}_2$. Since the particular $\boldsymbol{\beta}_2$ values of any active factors cannot be known a priori, we seek designs that minimize $tr(\boldsymbol{AA'})$ to minimize the impact of aliasing on the estimated model parameters.

The above criteria are two commonly used criteria for evaluating screening and supersaturated designs with different emphases ranging from the evaluation of orthogonality to the protection against model misspecification. There are many other criteria that could possibly be used with different design setups and goals. The methodology illustrated in this case study could be adapted for these alternative criteria.

Based on these two commonly used design criteria, we explore the catalog of all nonisomorphic designs in Johnson and Jones (2010) to suggest which subset of these represents "good" design choices to consider further and eliminate a large fraction of the alternatives as inferior. The Pareto front approach (Lu et al. 2011) is used to identify the set of contending designs and allows flexible exploration of trade-offs and balancing of priorities when combining multiple objectives.

The Pareto optimization has been broadly used in applications in different disciplines before being introduced as a structured decision-making process in the design of experiment paradigm (Lu et al. 2011, 2012). The method consists of two stages: (1) objective Pareto optimization, which assembles a set of superior designs as contenders by removing inferior choices from further consideration and (2) subjective Pareto decision analysis, to compare candidate designs by evaluating individual performance, trade-offs, and robustness to a spectrum of different emphases of the criteria using a set of graphical methods. This second stage concludes with a final decision, which chooses a best design based on priorities of the study. By separating the objective and subjective steps, an experimenter can first see the complete set of choices, before imposing any subjective experiment-specific considerations. By understanding the range of options and potential impacts of subjective choices, the decision maker is positioned to make an informed and defensible choice.

The first objective optimization stage finds the set of designs that are not strictly outperformed by any other designs in the entire design space. Here one design strictly outperforms another, or *Pareto dominates* it (in the terminology of the Pareto literature), if it is at least as good as another for all criteria and strictly better for at least one of the criteria. A *Pareto front* is formed in the criterion space with all designs that are not Pareto dominated by others. The Pareto set of designs represents an objective collection of options to select from, since for any alternative not on the Pareto front, there is at least one clearly better choice on the Pareto front. Hence, as the logical first step, finding the Pareto front allows one to see the complete set of superior options before making a subjective decision specific to a particular experimental scenario.

In the second stage, designs on the Pareto front are evaluated on three aspects with the graphics developed in Lu et al. (2011) and Lu and Anderson-Cook (2012): (1) finding the best solution for a particular weight combination that matches the user's study goals (from a particular location in the mixture plot in Lu et al. 2011), (2) the robustness of a chosen solution based on a range of weightings close to user preferences (the area of weightings where a solution is best in the mixture plot), and (3) individual design performance relative to the best available solution for a particular set of weight choices (with the synthesized efficiency plot from Lu and Anderson-Cook 2012). These graphical summaries allow quantitative evaluation of design choices from the Pareto front to be visualized for more intuitive comparison and matched with subjective choices affecting the final decision. As an essential part of the Pareto front approach, the graphical tools are helpful for making an informed, quantitatively based decision and for reaching consensus when there may be competing priorities for the study.

In the following section, we examine the 27 nonisomorphic regular and nonregular 16-run designs for six factors detailed in Johnson and Jones (2010) to categorize the alternatives based on the two criteria of $E(s^2)$ and $tr(\boldsymbol{AA}')$ and provide discussion about how to make a sensible and justifiable choice of a best design based on experimenter priorities. The next sections repeat the Pareto optimization process based on $E(s^2)$ and $tr(\boldsymbol{AA}')$ for the seven- and eight-factor 16-run designs with 55 and 80 candidate designs, respectively. Finally, the conclusions section highlights some of the key results for the different scenarios and discusses some options for expanding the set of criteria when making the decision and how this will impact the results of the Pareto front approach.

16-RUN SCREENING DESIGNS FOR SIX FACTORS

Johnson and Jones (2010, Appendix A) provide the design generating equations for the 27 nonisomorphic

16-run, six-factor, two-level regular and nonregular designs provided by Sun et al. (2002). The designs are categorized into four structural groups based on their aliasing patterns:

1. *Classical* designs have an embedded 2^4 full-factorial design for four of the factors and use defining equations that completely confound the other two main effects with sets of four two-, three-, or four-way interactions.
2. *Hybrid* designs have an embedded 2^4 full-factorial design with one other main effect completely confounded with four two-, three-, or four-way interactions and the other main effect partially confounded with four two-, three-, or four-way interactions.
3. *Correlated* designs have an embedded 2^4 full-factorial design and defining equations that partially confound the other two main effects with sets of four two-, three-, or four-way interactions.
4. *Replicated* designs use a replicated 2^3 factorial design as the starting point of design generation.

The values of the two criteria, $E(s^2)$ and $tr(\boldsymbol{AA}')$, as well as the category of their design structure for the 27 designs are shown in Table 1. The design index numbers are consistent with the numbering scheme in Johnson and Jones (2010). A scatterplot of the design criteria for the 27 designs is shown in Figure 1 with different symbols to distinguish between design structural groups. The nondominated designs on the Pareto front are highlighted with the dotted line.

Several designs have identical values for both criteria and hence are labeled with only one representative index number. Among the 27 designs, only five are on the Pareto front (shown in bold in Table 1). Designs 4, 13, and 14 are tied and optimal for $E(s^2)$. Design 5 is best for $tr(\boldsymbol{AA}')$. Design 8 represents a compromise with more balanced performance for the two criteria. The design generating equations for the five designs on the Pareto front are shown in Table 2.

The Utopia point, which corresponds to best values for both criteria and is unattainable for any design in this class, is shown with the solid circle at the bottom left corner of the plot. The Utopia point is typically identified as the "ideal" solution with the best available values for each of the criteria.

TABLE 1 **Values of the Two Criteria, $E(s^2)$ and $tr(AA')$, for 27 Nonisomorphic Six-Factor Designs and Their Corresponding Design Generating Structural Categories**[a]

Design no.	Category	$E(s^2)$	$tr(AA')$
1	Replicated	25.6	12
2	Classical	10.97	6
3	Classical	7.31	6
4	**Classical**	**7.31**	**3**
5	**Classical**	**10.97**	**0**
6	Hybrid	7.31	3.75
7	Hybrid	9.14	4.5
8	**Hybrid**	**9.14**	**1.5**
9	Hybrid	9.14	5.25
10	Replicated	18.29	9
11	Hybrid	10.97	6
12	Hybrid	10.97	3
13	**Hybrid**	**7.31**	**3**
14	**Hybrid**	**7.31**	**3**
15	Hybrid	9.14	4.5
16	Replicated	18.29	6
17	Correlated	10.97	6
18	Correlated	7.31	6
19	Correlated	9.14	3
20	Correlated	9.14	4.5
21	Replicated	14.63	7.5
22	Correlated	10.97	4.5
23	Correlated	9.14	5.25
24	Correlated	9.14	4.5
25	Replicated	14.63	6
26	Correlated	10.97	6
27	Replicated	12.8	6

[a]Designs on the Pareto front are in bold, among which designs 4, 13, and 14 are tied in both criteria.

However, this solution is typically not attainable because there is rarely an overall global winner, but the Utopia point is useful to serve as the gold standard when we evaluate the individual designs on the front based on how close they are located relative to the ideal solution. The Pareto front is located on the edge of the solution space closest to the Utopia point. By identifying the Pareto front, we automatically eliminate a large proportion of the design options in the candidate set because of their inferior performance relative to those on the Pareto front, and we can focus our attention on the most promising choices. For any design not on the Pareto front, one or more of designs 4, 5, 8, 13, or 14 is at least as good on each criterion and strictly better for at least one. Focusing attention on just those solutions on the Pareto front substantially

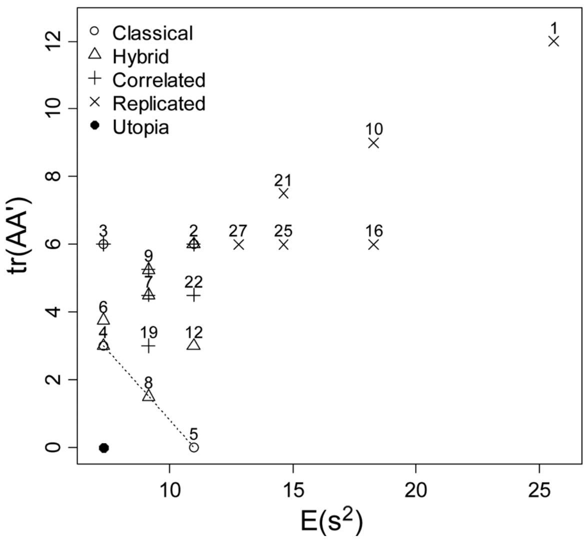

FIGURE 1 Scatterplot for 27 nonisomorphic designs with six factors based on the two criteria, $E(s^2)$ and $tr(AA')$. Designs on the Pareto front are connected with the dotted line. Designs from different categories (classical, hybrid, correlated, and replicated) are shown with different symbols. Designs that are tied in both criteria values (designs 2, 11, 17, and 26; designs 3 and 18; designs 4, 13, and 14; designs 7, 15, and 20; designs 9 and 23) are labeled with one index number as a representative. The Utopia point, which is the ideal solution with best values for both criteria, is shown with the solid dot at the bottom left corner.

reduces the time and effort needed for quantitative evaluation and comparison of the competitive designs for making a rational final decision.

Among the four categories of design generating structures, the replicated designs are all scattered in the top right area of Figure 1 and hence have inferior performance (large values for both criteria) among all 27 designs. The other three categories generally have relatively smaller $E(s^2)$ values and moderately small (no more than 50% of the maximum) $tr(\boldsymbol{AA}')$ values. However, no correlated designs are located

TABLE 2 Design Generating Equations for the Five 16-run, Six-Factor designs on the Pareto Front

Design no.	Category	Design generating equations for factors E and F (factors A–D are determined by a 2^4 factorial design)
4	Classical	E = AB, F = ACD
5	Classical	E = ABC, F = ABD
8	Hybrid	E = ABC, F = 1/2[CD + ACD + BCD − ABCD]
13	Hybrid	E = ABCD, F = 1/2[BD + ABD + CD − ACD]
14	Hybrid	E = ABC, F = 1/2[AD + BD + ABCD − CD]

on the Pareto front, which indicates that none of them are optimal regardless of how the two criteria are valued. Designs 4 and 5, which are optimal based on only $E(s^2)$ or $tr(\boldsymbol{AA}')$, respectively, are both classical designs. The remaining three designs (8, 13, and 14) on the front are hybrid designs.

The identification of the Pareto front has narrowed the design choices to only three sensible options (five designs including the ties) based on the criteria $E(s^2)$ and $tr(\boldsymbol{AA}')$. However, to actually conduct the experiment, the practitioner has to ultimately choose only a single design to run. Hence, we evaluate the individual design performance, trade-offs, and robustness to different emphases of the relative importance of the two criteria using the graphical tools developed in Lu et al. (2011), Lu and Anderson-Cook (2012), and Lu, Chapman, and Anderson-Cook (2013).

The Utopia point approach (Lu et al. 2011) is used to further select optimal designs from the Pareto front based on a user specified desirability function (Derringer and Suich 1980) for combining the two criteria. A fine grid of weight combinations spreading across the entire possible weighting space (from 100% of the weight for $E(s^2)$ to 100% of the weight for $tr(\boldsymbol{AA}')$) is evaluated to explore which design is best for different subjective weighting choices. To use the desirability function approach, both criteria values are converted to a 0–1 scale with a linear transformation by matching the worst and best desirable values to 0 and 1, respectively. For example, if the worst and best desirable values for $tr(\boldsymbol{AA}')$ are 12 and 0 for an experiment, then a design with $tr(\boldsymbol{AA}')$ equaling 6 will have a value of 0.5 on the converted 0–1 desirability scale. The desirability value of 1 for each criterion typically corresponds to the optimal value observed. However, there could be alternative ways for choosing the scale for the worst desirability value (an admitted oxymoron). Choosing the worst-performing design among those on the Pareto front can be used when the candidate set of designs is extremely large (impractical to consider all possibilities). Alternatively, if the entire population of candidate choices can be considered (e.g., the 27 designs for the six-factor case), set the value of 0 for the worst-performing design in the population. A final possibility is to apply a user-specified value reflecting subject-matter knowledge. Besides the scaling scheme, a metric is needed to integrate multiple criteria into a single summary index for ranking the design performance. A multiplicative desirability function for design j in the form of

$$DF(j, \boldsymbol{w}) = s_1^{w_1}(j) s_2^{w_2}(j), \qquad [4]$$

where $\boldsymbol{w} = (w_1, \ w_2)$ with $0 \le w_1, \ w_2 \le 1$ and $w_1 + w_2 = 1$, and s_1 and s_2 are the scaled criteria values between 0 and 1, is a common choice, which penalizes poor design performance for at least one of the criteria rather severely. This is equivalent to using the log L_1-norm metric in the Utopia point approach literature (Lu et al. 2011). Another common choice is the additive desirability function, which combines multiple criteria as a weighted sum of the scaled criteria values and allows superior performance of one criterion to overcome the poor performance of another criterion. What desirability function (DF) form to choose depends on whether the experimenter wants severe penalization for poor performance of a certain criterion (multiplicative DF) or whether stellar performance in one criterion is thought to be a reasonable trade-off for very poor performance on the other (additive DF). Note that the optimal design that is selected depends on the choice of scaling scheme and desirability function form. Different choices can be used for different user preferences and priorities. The Pareto front approach offers considerable flexibility to explore different choices and conduct a sensitivity analysis with little extra computational effort, especially for scenarios with a large number of possible candidates (Lu et al. 2011).

Next we explore the potential impact of these subjective choices on decision making. First consider a scenario of using the Pareto front scaling (worst value of designs on the Pareto front is mapped to 0) combined with the multiplicative desirability function. The design that maximizes [4] is the optimal design for a particular weighting choice, $\boldsymbol{w}$. The optimal designs for all possible different weighting choices are shown in the mixture plot in Figure 2a. This was adapted from the mixture plot developed for mixture designed experiments in Cornell (2002). For the three-criteria case, the mixture plot is a triangular simplex. For the two-criteria scenario, the plot collapses to a horizontal line segment. Moving from left to right in Figure 2a, the relative priorities shift from weighting $tr(\boldsymbol{AA}')$ heavily to more emphasis on $E(s^2)$. Since the two weights sum to 1, knowing the weight of one of the criteria specifies the other weight completely. The mixture plot can also be

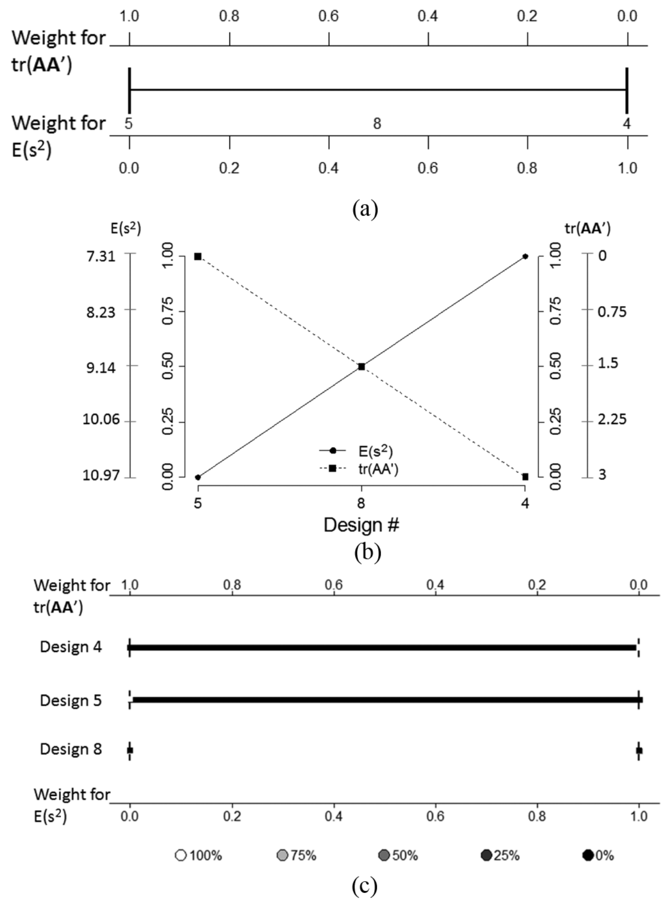

FIGURE 2 Graphical summaries for selecting six-factor designs from the Pareto front based on using the multiplicative desirability function based on the criteria values of designs on the Pareto front: (a) mixture plot for showing the optimal designs for different weightings of the two criteria; (b) trade-off plot for optimal designs selected in the mixture plot; and (c) synthesized efficiency plot for optimal designs selected in the mixture plot.

adapted for the four criteria scenario, in which case the range of all combinations of weighting preferences can be displayed in a tetrahedron (Lu and Anderson-Cook 2014). Design 8 is optimal for all possible weights in (0, 1) except for the two extreme weightings with 100% weight for only one of the criteria. Design 4 (as well as the tied designs 13 and 14) is optimal only when $E(s^2)$ is weighted 100%. Design 5 is the best design only if $tr(\boldsymbol{AA}')$ is given a 100% weight.

The mixture plot provides a mechanism to align the study goals with the experimenter's particular region of interest via the weighting. For example, in our case study, if the experimenter has about equal concern on both criteria, then the focus would be the region around the middle area of mixture plot. Depending on how much uncertainty is associated with this preference, the weight region could be either as narrow as allowing 45%–55% weight for each criterion or it could be as wide as allowing any weighting between 30% and 70% being possible for each criterion. With the experimenter's weighting region of interest, the mixture plot leads to the possible optimal designs to consider in that region. However, for our particular case, design 8 is the absolute dominating choice for all weighting preferences except for the extreme cases with 100% interest in a single criterion.

Figure 2b shows the trade-offs between the different designs selected from the Pareto front. The inner axes of the plot are generated based on the desirability (0–1) scale, with the raw criteria scales shown on both sides of the plot. Designs are sorted by worst to best $E(s^2)$ values from left to right. The simple trade-off pattern observed for the three optimal designs (ignoring the ties) captures the simple, nearly linear shape of the Pareto front. Designs 4 and 5 are both 100% desirable for one of the two criteria, and design 8 has equally balanced performance (50% desirable) for both criteria.

Figure 2c shows the synthesized efficiency plots (Lu and Anderson-Cook 2012) for the three designs. It shows the individual design performance based on quantifying its performance relative to the optimal for a spectrum of different weighting choices. The relative design performance for design j is quantified by its synthesized efficiency as a function of a particular set of weights, $\boldsymbol{w}$, defined as

$$SE(j, \boldsymbol{w}) = DF(j, \boldsymbol{w}) / max_j\{DF(j, \boldsymbol{w})\}.$$

To calculate the synthesized efficiency values for a particular solution, the DF is valued for every combination of weights and then compared to the best DF value at that weight. High to low synthesized efficiency values are plotted with a white–gray–black scale with 20 shades of gray each corresponding to a 5% band of efficiency values. Hence, designs 4 and 5 are black for the entire weight interval except for one extreme end, and design 8 is white (and optimal) for all weights except the endpoint cases.

Hence, when a multiplicative DF is preferred and the criteria values are scaled based on values from the Pareto front, the hybrid design 8 is the dominant choice. Designs 4 (13 and 14) and 5 are optimal when only a single criterion is considered. On the other hand, if the experimenter prefers the additive DF for combining the criteria, then design 8 is optimal for only a single weight combination with equal weight for both criteria, and designs 4 (13 and 14) and 5 would be selected based on whether $E(s^2)$ or $tr(\boldsymbol{AA}')$ is valued as more important among the two criteria.

Since we have only a fixed finite set of candidate designs to choose from, this naturally defines the complete range of values in this space for each of the criteria. We explore another scaling based on values from all designs in the population following the same process for using the scaling based on only designs on the Pareto front. Suppose the multiplicative DF in [4] is selected for combining the criteria, then the optimal designs for different weighting choices are summarized in the mixture plot shown in Figure 3a. With this alternative scaling, design 8

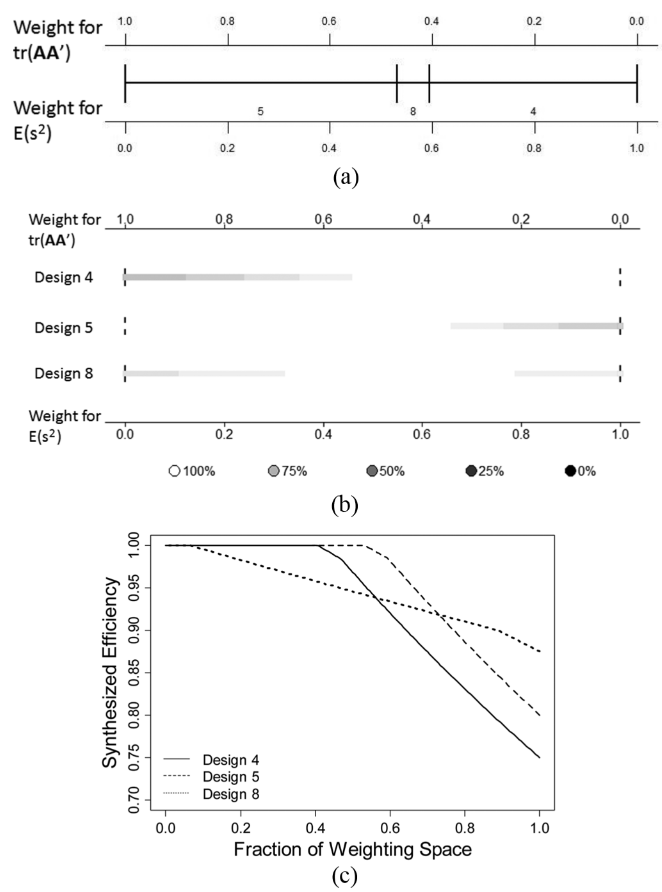

FIGURE 3 Graphical summaries for selecting six-factor designs from the Pareto front based on using the multiplicative desirability function with an alternative scaling based on the criteria values of the population of candidate choices: (a) mixture plot for showing the optimal designs for different weightings of the two criteria; (b) synthesized efficiency plot for optimal designs selected in the mixture plot; and (c) FWS plot for the selected designs.

is no longer the dominant choice but is only optimal when $E(s^2)$ is weighted between about 53% and 60%. In contrast, design 5, which under the previous scaling was optimal when only $tr(\boldsymbol{AA}')$ criterion is considered, is now optimal for around 53% of the possible weightings when $tr(\boldsymbol{AA}')$ is weighted between 47% and 100%. When $E(s^2)$ is weighted at least 60% of the weight, design 4 (tied with designs 13 and 14) is the best choice.

Figure 3b shows the synthesized efficiency plot for designs 4, 5, and 8. All three designs have no less than a 75% synthesized efficiency for all possible weightings. Design 4 has a large white region (corresponding to at least 95% efficient) when more weight is given to $E(s^2)$, and design 5 is above 95% efficient for 60% of the weighting space. Design 8 has the shortest white region but has the largest minimum synthesized efficiency (corresponding to lightest dark color across the weighting space). The differences between the results from Figures 2 and 3 are due to the large difference in the range of values for the Pareto front, and the entire population of designs leads to dramatic differences in the scaled values with the two different scaling schemes. The Pareto front scaling results in much bigger (almost five times) trade-offs between designs than the population scaling and hence requires more balancing between the two criteria. The substantial changes in the results of using different scaling schemes (Figure 2 vs. Figure 3) indicate how big an impact the subjective choice can have on the solution. Hence, it is advantageous to use the Pareto front approach since it is computationally more efficient to conduct a sensitivity analysis of the subjective factors based on evaluating only a smaller set of choices on the Pareto front.

To summarize the individual design performance across the entire weighting space, Figure 3c shows the fraction of weighting space (FWS) plot (Lu, Chapman, and Anderson-Cook 2013) for the three designs. The line for each design displays the fraction of the weighting space where the design has synthesized efficiency at least as high as the specified percentage. This provides an overall quantitative summary of individual design performance across the entire weighting space and hence allows for an easy and intuitive comparison of several design choices when all of the weights are considered of interest. A discrete approximation of this summary can be built from the information in the synthesized efficiency plot in Figure 3b. For each design, we have a vector of synthesized efficiency values relative to the best possible for each weight combination with fine coverage of the entire weighting space. Then we sort the efficiency values in descending order and extract a list of distinct values in the same order. For each distinct value, we calculate the fraction of entries (weight combinations) in the sorted efficiency vector at least as large as that value. This graphical summary is implemented in R, with scripts available from the first author upon request. Design 8 has the best minimum synthesized efficiency of 87.5%. Designs 4 and 5 have higher synthesized efficiencies than design 8 for close to 47% and 60% of weighting space, respectively. However, their synthesized efficiencies drop much faster after these initial high values with the minimum values at 75% and 80% for designs 4 and 5, respectively. If all possible weightings are considered of equal interest, design 5 has generally better performance with consistently higher synthesized efficiency for around 60% of the weighting space. However, if there is a more focused region of interest for how to weight the two criteria, then different solutions may be selected depending on where the experimenters' priorities lie, how big the weighting region of interest is, and whether the average or the worst case of performance is more of interest. Lu, Anderson-Cook, and Lin (2013) adapted the FWS plot for flexibly incorporating more focused weighting preference for two criteria when summarizing across only a portion of the range of interest.

The additive desirability function is also examined based on the scaling from the entire population of designs, with design 5 chosen as optimal when $E(s^2)$ is weighted less than 56% and design 4 is best for the remaining weightings. Design 8 is not optimal for any weight combinations. Alternate scaling or DF forms would also be possible and should be chosen to match experimenter goals. Adjusting the analyses based on different choices is straightforward given the table of values in Table 1. Regardless of what scaling or desirability function forms the experimenter chooses, the identification of the Pareto front is independent of those choices. The exploration of different possible scenarios with different weighting, scaling, and DF choices indicates that our solution is dependent on these subjective choices. However, the selection of the Pareto front substantially reduces the set of promising choices to consider further (from 27 designs down to 5) and allows for more efficient evaluation of the sensitivity of solutions to different subjective weighting and scaling choices. The overall message from the analysis is that one of designs 4, 5, 8, 13, or 14 should be chosen as an ideal solution based on $E(s^2)$ and $tr(\boldsymbol{AA'})$. Which one of these is best for a particular experiment is based on how the experimenter values the criteria and how much penalty to assign to inferior performance.

16-RUN SCREENING DESIGNS FOR SEVEN FACTORS

This section considers selecting 16-run two-level screening designs with seven factors from the catalog of 55 nonisomorphic regular and nonregular designs based on the two general criteria, $E(s^2)$ and $tr(\boldsymbol{AA'})$. Table 3 contains the criteria values and corresponding design structural categories for the 55 designs with the same numbering scheme as in Johnson and Jones (2010, Appendix B). A scatterplot of the 55 designs is shown in Figure 4 with different symbols for different categories. Designs with identical criteria values are labeled with only one representative index number, and details of the ties are listed in the caption. Designs on the Pareto front are connected with the dotted line, which are on the edge of the population closest to the unattainable Utopia point at the bottom left corner.

There are 10 designs on the Pareto front with four groups of distinctive criteria value pairs. Design 5, which is tied with designs 11, 21, 22, 32, and 33, is optimal for $E(s^2)$. Design 6 has the best $tr(\boldsymbol{AA'})$ value. Design 26 (tied with design 28) and design 12 have moderate values of the two criteria and represent compromise choices. The 10 designs on the Pareto front with their criterion values are shown in bold in Table 3 with their corresponding design generating equations (Johnson and Jones 2010) contained in Table 4. Similar to the six-factor case, all of the replicated designs have relatively poor performance for both criteria. Among the 10 designs on the Pareto front, two are classical, seven are hybrid, and only one is a correlated design. The two classical designs 5 and 6 are optimal for one of the two criteria with substantial sacrifice for the other criterion. Several

TABLE 3 Values of Criteria, $E(s^2)$ and $tr(AA')$, for the 55 Nonisomorphic Seven-Factor Designs and Their Corresponding Design Structural Categories[a]

Design no.	Category	$E(s^2)$	$tr(\boldsymbol{AA'})$
1	Replicated	28.44	21
2	Classical	14.22	12
3	Classical	12.19	9
4	Classical	10.16	9
5	**Classical**	**10.16**	**6**
6	**Classical**	**14.22**	**0**
7	Hybrid	10.16	7.5
8	Hybrid	10.67	8.25
9	Hybrid	12.19	9
10	Hybrid	10.16	9
11	**Hybrid**	**10.16**	**6**
12	**Hybrid**	**12.19**	**3**
13	Hybrid	10.16	9
14	Hybrid	12.19	10.5
15	Hybrid	10.16	7.5
16	Replicated	20.32	15
17	Hybrid	14.22	12
18	Hybrid	12.19	9
19	Hybrid	10.16	9
20	Hybrid	14.22	6
21	**Hybrid**	**10.16**	**6**
22	**Hybrid**	**10.16**	**6**
23	Hybrid	10.67	6.75
24	Hybrid	12.19	9
25	Hybrid	10.16	9
26	**Hybrid**	**11.17**	**4.5**
27	Hybrid	11.17	7.5
28	**Hybrid**	**11.17**	**4.5**
29	Hybrid	10.67	8.25
30	Hybrid	10.16	7.5
31	Correlated	12.19	6
32	**Correlated**	**10.16**	**6**
33	**Hybrid**	**10.16**	**6**
34	Hybrid	11.17	9
35	Hybrid	10.67	8.25
36	Hybrid	10.67	6.75
37	Hybrid	10.16	7.5
38	Correlated	12.19	9
39	Correlated	10.16	9
40	Hybrid	12.19	10.5
41	Replicated	16.25	12
42	Replicated	14.22	12
43	Hybrid	10.16	7.5
44	Correlated	14.22	9
45	Correlated	12.19	9
46	Correlated	10.16	7.5
47	Correlated	10.16	9
48	Correlated	11.17	9
49	Correlated	11.17	6
50	Correlated	11.17	9
51	Replicated	14.22	10.5
52	Replicated	14.22	10.5
53	Correlated	10.16	7.5

(*Continued*)

TABLE 3 Continued

Design no.	Category	$E(s^2)$	$tr(AA')$
54	Correlated	10.67	8.25
55	Correlated	10.67	6.75

[a]Designs on the Pareto front are shown in bold, among which design 5 is tied with designs 11, 21, 22, 32, and 33, and design 26 is tied with design 28 in both criteria.

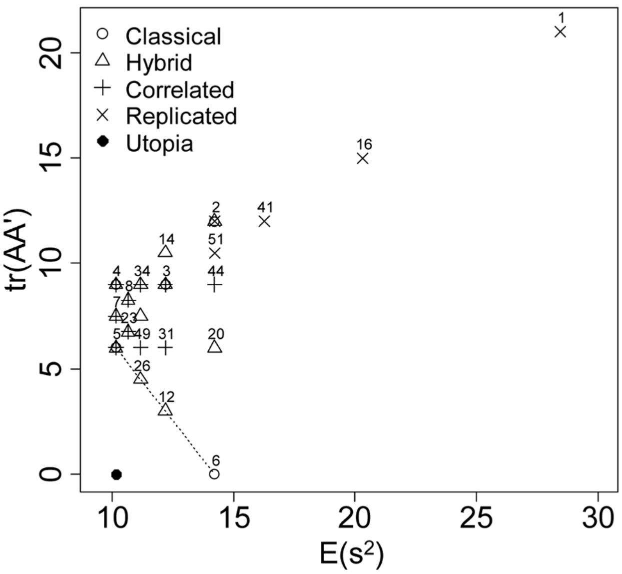

FIGURE 4 **Scatterplot for 55 nonisomorphic 16-run designs with seven factors based on the two criteria, $E(s^2)$ and $tr(AA')$. Designs from different categories (classical, hybrid, correlated, and replicated) are shown with different symbols. Designs that are tied in both criteria values are labeled with one index number as a representative. Designs on the Pareto front are connected with the dotted line (among which design 5 is tied with designs 11, 21, 22, 32, and 33 and design 26 is tied with design 28). The Utopia point, which is the ideal solution with best values for both criteria, is shown with the solid dot at the bottom left corner.**

of the hybrid designs (designs 12, 26, and 28) are more balanced between the two criteria.

The Pareto front eliminates more than 80% of all possible designs in the objective stage and allows a final decision to be made from only the four distinct criterion value options. Next, we compare the remaining designs based on their trade-offs and robustness to different weighting choices using the same set of graphical summaries as for the six-factor case. We illustrate the case with the subjective choice of the multiplicative desirability function combined with the scaling based on the range of criteria values for the entire population of 55 designs. However, the experimenter has the flexibility of exploring and choosing different scaling and desirability functions based on their preferences/priorities.

Figure 5a shows the mixture plot of the optimal designs for different weighting choices. Design 6 is optimal for about 53% of possible weights when $tr(\boldsymbol{AA}')$ is valued more. Design 5 (tied with designs 11, 21, 22, 32, and 33) is optimal when $E(s^2)$ is weighted at least 63%. Design 12 is best when $E(s^2)$ is weighted between 53% and 59%, and design 26 (tied with design 28) is the optimal solution when $E(s^2)$ is weighted between 59% and 63%. Figure 5b shows the trade-off plot for the four designs from the mixture plot. All four designs achieve as least 70% of the best performance for both criteria based on the chosen scaling. Designs 5 and 6 have the most trade-off (100% best performance for one criterion and less than 80% for the other) among designs on the front, and design 12 is most balanced with 85%–90% of best performance for both criteria.

TABLE 4 **Design Generating Equations for the Ten 16-run, Seven-Factor Designs on the Pareto Front**[a]

Design #	Category	Design generating equations for factors E, F, and G (factors A–D are determined by a 2^4 factorial design)
5	Classical	E = AB, F = AC, G = BCD
6	Classical	E = ABC, F = ABD, G = ACD
11	Hybrid	E = BD, F = ACD, G = 1/2[ABC + ABD + ABCD − AB]
12	Hybrid	E = ABC, F = ABD, G = 1/2[CD + ACD + BCD − ABCD]
21	Hybrid	E = BCD, F = 1/2[BD + ABD + CD − ACD], G = 1/2[ABD + CD + ACD − BD]
22	Hybrid	E = ABD, F = ABC, G = 1/2[AD + BD + CD − ABCD]
26	Hybrid	E = ABC, F = 1/2[BD + ABD + BCD − ABCD], G = 1/2[BD + CD + ACD − ABD]
28	Hybrid	E = BCD, F = 1/2[AC + ACD + ABC − ABCD], G = 1/2[AC + ACD + AB − ABD]
32	Correlated	E = 1/2[AC + ABC + ACD − ABCD], F = 1/2[AB + AC + ABD − ACD], G = 1/2[ABC + ABD + ABCD − AB]
33	Hybrid	E = BCD, F = 1/2[AC + AD + ABD − ABC], G = 1/2[AB + AD + ABCD − AC]

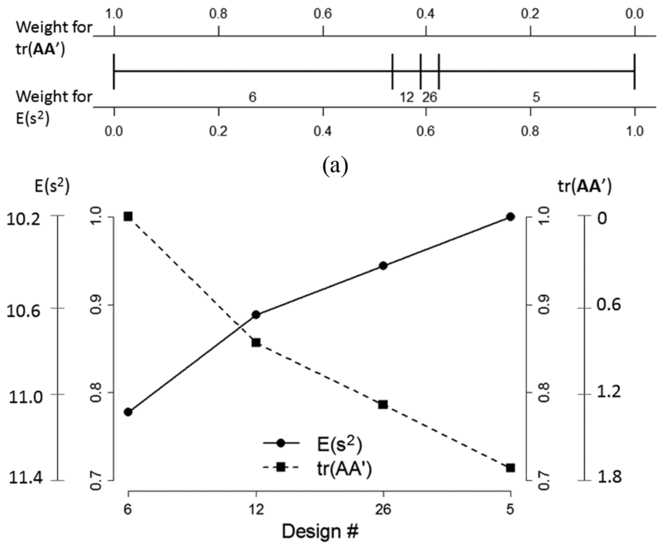

FIGURE 5 (a) Mixture plot and (b) trade-off plot for selecting seven-factor designs from the Pareto front with the multiplicative desirability function and the scaling based on the population of candidate choices. Designs 6 and 5 are both optimal for one of the two criteria but have relatively poor performance for the other criterion and are quite robust optimal choices when one criterion is valued substantially more important than the other criterion. Designs 12 and 26 represent compromise choices with more balanced performance between the two criteria. However, they are optimal for only a small region of weightings and hence have limited robustness to weight uncertainty.

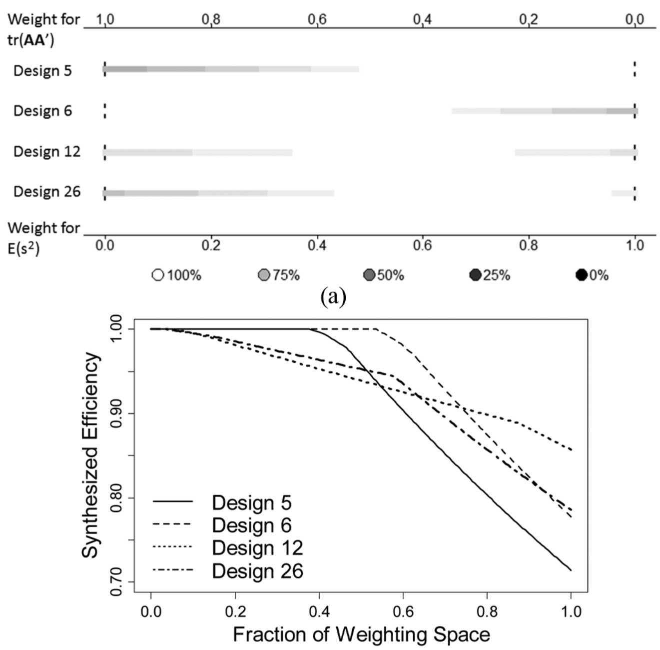

FIGURE 6 (a) Synthesized efficiency plot and (b) FWS plot for selecting seven-factor designs from the Pareto front with the multiplicative desirability function and the scaling based on the population of candidate choices.

Individual design performance relative to the optimal for different weighting choices is shown in the synthesized efficiency plots in Figure 6a. Design 6 has the largest white region with at least 95% synthesized efficiency when $E(s^2)$ is weighted less than 66%. Design 5 is at least 95% efficient when $E(s^2)$ is weighted at least 51%. The worst synthesized efficiency values for designs 6 and 5 are around 77% and 70%, respectively, when considering only one of the criteria. Design 26 has good performance denoted by the white region for around 52% of the weighting space with minimum synthesized efficiency of 78%. Design 12 is at least 95% efficient for 42% of possible weightings and has the largest minimum efficiency around 86%. By summarizing across the entire weighting space, Figure 6b shows the FWS plot for the four designs. Design 12 has the flattest curve with best (largest) minimum synthesized efficiency; however, it has the narrowest high efficiency region (synthesized efficiency above 95%). Design 6 has the largest weighting space with at least 90% synthesized efficiency, with the lower end of efficiency dropping quickly for the worst 25% of the weighting space.

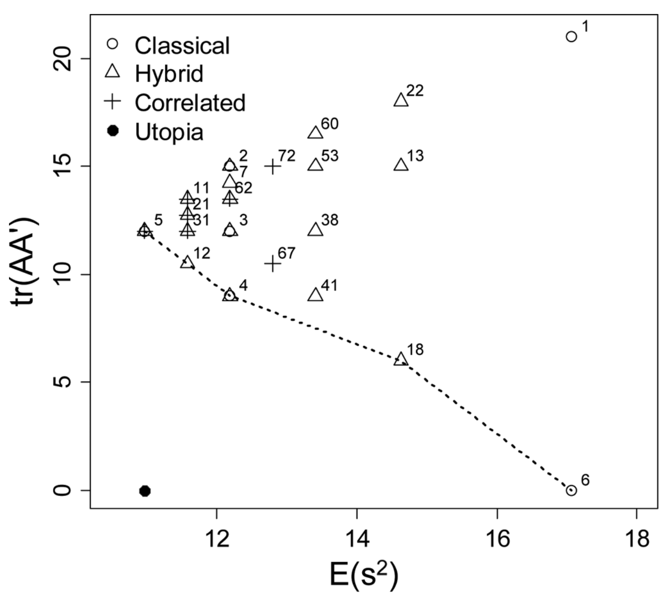

FIGURE 7 Scatterplot for 80 nonisomorphic 16-run designs with eight factors based on the two criteria, $E(s^2)$ and $tr(AA')$. Designs from different categories (classical, hybrid, and correlated) are shown with different symbols. Designs that are tied in both criteria values are labeled with one index number as a representative. Designs on the Pareto front are connected with the dotted line (among which design 4 is tied with designs 17, 26, 42, 48, and 77; design 5 is tied with designs 9, 16, 20, 25, 28, 30, 36, 50, 61, and 63; and design 12 is tied with designs 40, 47, 58, and 76). The Utopia point, which is the ideal solution with best values for both criteria, is shown with the solid dot at the bottom left corner.

TABLE 5 **Values of Criria, $E(s^2)$ and $tr(AA')$, for the 80 Nonisomorphic Eight-factor Designs and their Corresponding Design Structural Categories**

Design no.	Category	$E(s^2)$	$tr(AA')$
1	Classical	17.07	21
2	Classical	12.19	15
3	Classical	12.19	12
4	**Classical**	**12.19**	**9**
5	**Classical**	**10.97**	**12**
6	**Classical**	**17.07**	**0**
7	Hybrid	12.19	14.25
8	Hybrid	12.19	12
9	**Hybrid**	**10.97**	**12**
10	Hybrid	12.19	15
11	Hybrid	11.58	13.5
12	**Hybrid**	**11.58**	**10.5**
13	Hybrid	14.63	15
14	Hybrid	12.19	15
15	Hybrid	12.19	12
16	**Hybrid**	**10.97**	**12**
17	**Hybrid**	**12.19**	**9**
18	**Hybrid**	**14.63**	**6**
19	Hybrid	12.19	15
20	**Hybrid**	**10.97**	**12**
21	Hybrid	11.58	12.75
22	Hybrid	14.63	18
23	Hybrid	12.19	15
24	Hybrid	12.19	12
25	**Hybrid**	**10.97**	**12**
26	**Hybrid**	**12.19**	**9**
27	Hybrid	12.19	12
28	**Hybrid**	**10.97**	**12**
29	Hybrid	11.58	13.5
30	**Hybrid**	**10.97**	**12**
31	Hybrid	11.58	12
32	Hybrid	12.19	14.25
33	Hybrid	11.58	12.75
34	Hybrid	11.58	13.5
35	Hybrid	12.19	12
36	**Hybrid**	**10.97**	**12**
37	Hybrid	11.58	13.5
38	Hybrid	13.41	12
39	Hybrid	12.19	12
40	**Hybrid**	**11.58**	**10.5**
41	Hybrid	13.41	9
42	**Hybrid**	**12.19**	**9**
43	Hybrid	11.58	13.5
44	Hybrid	12.19	15
45	Hybrid	11.58	12
46	Hybrid	11.58	13.5
47	**Hybrid**	**11.58**	**10.5**
48	**Hybrid**	**12.19**	**9**
49	Hybrid	12.19	12
50	**Hybrid**	**10.97**	**12**
51	Hybrid	11.58	13.5
52	Hybrid	11.58	12.75
53	Hybrid	13.41	15
54	Hybrid	12.19	15
55	Hybrid	12.19	12
56	Hybrid	11.58	13.5
57	Hybrid	11.58	13.5
58	**Hybrid**	**11.58**	**10.5**
59	Hybrid	12.19	15
60	Hybrid	13.41	16.5
61	**Hybrid**	**10.97**	**12**
62	Hybrid	12.19	13.5
63	**Correlated**	**10.97**	**12**
64	Correlated	11.58	12.75
65	Correlated	11.58	13.5
66	Hybrid	11.58	12
67	Correlated	12.80	10.5
68	Correlated	12.80	10.5
69	Hybrid	11.58	13.5
70	Hybrid	11.58	13.5
71	Hybrid	11.58	12.75
72	Correlated	12.80	15
73	Hybrid	11.58	12.75
74	Hybrid	11.58	13.5
75	Hybrid	12.19	14.25
76	**Hybrid**	**11.58**	**10.5**
77	**Hybrid**	**12.19**	**9**
78	Correlated	11.58	13.5
79	Correlated	11.58	12
80	Correlated	12.19	13.5

[a]Designs on the Pareto front are shown in bold, among which design 5 is tied with designs 9, 16, 20, 25, 28, 30, 36, 50, 61, and 63, design 4 is tied with designs 17, 26, 42, 48, and 77, and design 12 is tied with designs 40, 47, 58, and 76 in both criteria.

With the above quantitative information for design evaluation and comparison, the final decision should be made based on where the experimenter's weighting preference lies, how much uncertainty there is associated with that choice of weighting range, as well as the experimenter's tolerance for poor performance. Again, the overall message from the analysis should be that one of 10 designs should be chosen as an ideal solution based on $E(s^2)$ and $tr(\boldsymbol{AA'})$. Which one is best for an experiment depends on the experimenter's priorities.

16-RUN SCREENING DESIGNS FOR EIGHT FACTORS

This section examines the eight-factor 16-run case based on evaluating the 80 nonisomorphic designs given in Johnson and Jones (2010, Appendix C). A scatterplot of the 80 designs is shown in Figure 7 with the criteria values given in Table 5. The 80 designs have only three categories of structures, with no replicated designs being generated from a 2^3 factorial starting point. The Pareto front identifies 24 nondominated designs (3 classical, 1 correlated, and 20 hybrid designs) with five groups of distinctive criteria values and eliminates 70% of designs for further consideration. Design 5, tied with designs 9, 16, 20, 25, 28, 30, 36, 50, 61, and 63, is optimal for $E(s^2)$. Design 6 is optimal for $tr(\boldsymbol{AA}')$. Design 12 (tied with 40, 47, 58, and 76), design 4 (tied with 17, 26, 42, 48, and 77), and design 18 are compromise choices with different degrees of balancing between the two criteria. The defining equations for all designs on the Pareto Front are given in Table 6.

Figures 8 and 9 show the graphical summaries based on one possible choice of subjective choices with the multiplicative DF and the population-based scaling. From Figures 8a and 8b, design 6 with the largest trade-off between two criteria is optimal only when $tr(\boldsymbol{AA}')$ criterion is given 100% of the weight. Designs 18 and 5 are best when $E(s^2)$ is weighted less than 24% and at least 5%, respectively. Design 4 is optimal when the weight for $E(s^2)$ is between 24% and 53% including the scenario when the two criteria are valued equally. Design 12 is optimal for around 6% of the remaining weighting area. Depending on what region of weights is of interest, a different

TABLE 6 **Design Generating Equations for the Twenty-Four, 16-run Eight-Factor Designs on the Pareto Front**

Design #	Category	Design generating equations for factors E, F, G, and H (factors A–D are determined by a 2^4 factorial design)
4	Classical	E = AB, F = AC, G = AD, H = BCD
5	Classical	E = AB, F = AC, G = BD, H = CD
6	Classical	E = ABC, G = ABD, G = ACD, H = BCD
9	Hybrid	E = CD, F = BD, G = AC, H = 1/2[AB + ABD + ABC − ABCD]
12	Hybrid	E = BD, F = BC, G = ACD, H = 1/2[ABC + ABD + ABCD − AB]
16	Hybrid	E = BD, F = AC, G = 1/2[AB + ABD + ABC − ABCD], H = 1/2[ABC + ABD + ABCD − AB]
17	Hybrid	E = BD, F = ACD, G = 1/2[AB + ABD + ABC − ABCD], H = 1/2[ABD + ABC + ABCD − AB]
18	Hybrid	E = BCD, F = ACD, G = 1/2[AB + ABD + ABC − ABCD], H = 1/2[ABD + ABC + ABCD − AB]
20	Hybrid	E = BD, F = CD, G = ABC, H = 1/2[AB + ABD + AC − ACD]
25	Hybrid	E = BD, F = ABC, G = 1/2[AB + ABD + AC − ACD], H = 1/2[ABD + AC + ACD − AB]
26	Hybrid	E = BCD, F = ABD, G = ACD, H = 1/2[AD + AB + AC − ABCD]
28	Hybrid	E = BCD, F = ABC, G = 1/2[AB + AC + AD − ABCD], H = 1/2[AB + AD + ABCD − AC]
30	Hybrid	E = CD, F = BD, G = 1/2[AC + ACD + ABC − ABCD], H = 1/2[AC + AB + ABD − ACD]
36	Hybrid	E = CD, F = 1/2[AC + ACD + ABC − ABCD], G = 1/2[AC + AB + ABD − ACD], H = 1/2[ABD + ABC + ABCD − AB]
40	Hybrid	E = BCD, F = AD, G = 1/2[AC + ACD + ABC − ABCD], H = 1/2[AC + AB + ABD − ACD]
42	Hybrid	E = BCD, F = 1/2[AC + ACD + ABC − ABCD], G = 1/2[AC + AB + ABD − ACD], H = 1/2[ABD + ABC + ABCD − AB]
47	Hybrid	E = BCD, F = ACD, G = 1/2[AC + AD + ABD − ABC], H = 1/2[AC + AB + ABCD − AD]
48	Hybrid	E = BCD, F = ACD, G = 1/2[AD + AC + ABC − ABD], H = 1/2[ABC + ABD + ABCD − AB]
50	Hybrid	E = BCD, F = 1/2[AC + AD + ABC − ABD], G = 1/2[AC + AB + ABCD − AD], H = 1/2[ABC + ABD + ABCD − AB]
58	Hybrid	E = ACD, F = 1/2[AC + AD + ABC − ABD], G = 1/2[AB + ABC + ABD − ABCD], H = 1/2[ABC + ABD + ABCD − AB]
61	Hybrid	E = ABCD, F = 1/2[AD + ABD + BCD − CD], G = 1/2[AD + CD + BCD − ABD], H = 1/2[AD + BD + ACD − BCD]
63	Correlated	E = 1/2[AC + ABC + AD − ABD], F = 1/2[AC + AD + ABD − ABC], G = 1/2[AC + BC + BD − AD], H = 1/2[BC + AD + BD − AC]
76	Hybrid	E = ABC, F = 1/2[AD + BD + CD − ABCD], G = 1/2[AC + ABC + ABCD − ACD], H = 1/2[AD + ACD + BCD − BD]
77	Hybrid	E = ABC, F = 1/2[AD + ABD + ABCD − ACD], G = 1/2[AD + ACD + BCD − BD], H = 1/2[ABD + CD + BCD − AD]

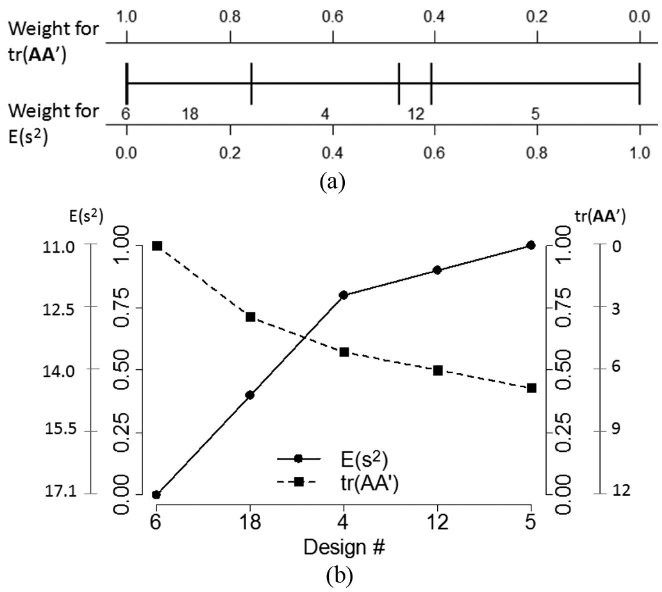

FIGURE 8 **(a) Mixture plot and (b) trade-off plot for selecting eight-factor designs from the Pareto front with the multiplicative desirability function and the scaling based on the population of candidate choices.**

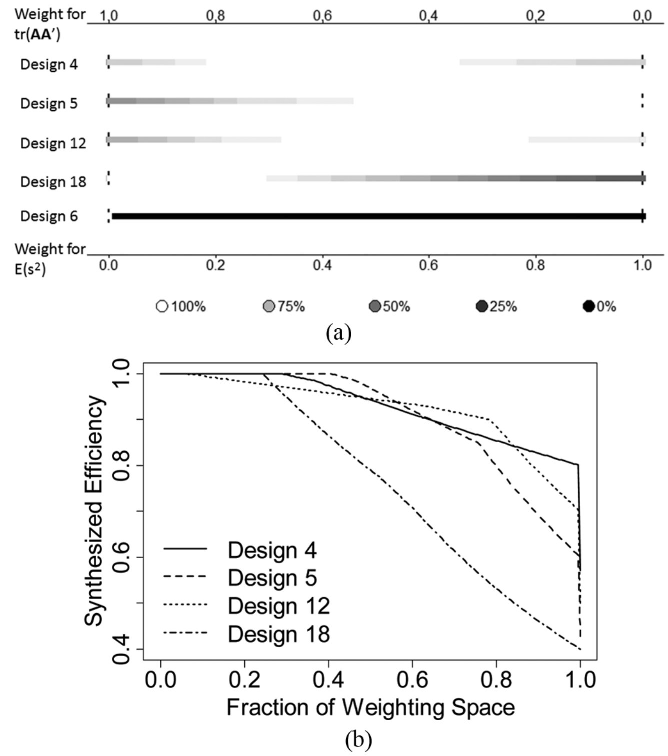

FIGURE 9 **(a) Synthesized efficiency plot and (b) FWS plot for selecting eight-factor designs from the Pareto front with the multiplicative desirability function and the scaling based on the population of candidate choices.**

optimal design will be selected. From Figures 9a and 9b, design 4 has the overall flattest FWS curve with at least 95% synthesized efficiency for about 47% of the weighting space and highest worst-case synthesized efficiency. Design 12 has a similar proportion of high synthesized efficiency area (≥95%) but with the worst-case values dropping most quickly. Design 5 has at least 95% synthesized efficiency for the largest (around 55%) weighting area among all designs on the Pareto front; however, it has the lowest synthesized efficiency for about 25% of the weights. Design 18 is at least 95% efficient when $E(s^2)$ is weighted less than 30% but has the lowest synthesized efficiency among all designs on the Pareto front for the remaining weighting space. Generally, designs 6 and 18 are less desirable. Design 4 is best if the entire weighting space is of interest. Design 5 or 12 should be selected depending on whether best performance in the high-efficiency region or low-efficiency region is considered more important.

Again, the overall message from the analysis is that only 24 designs should be considered as possible ideal solutions based on $E(s^2)$ and $tr(\mathbf{AA'})$. Which one is best for an experiment depends on the experimenter's priorities. The same process is followed for conducting the analysis. However, the results are helpful for demonstrating how decisions are made for different complication levels of the Pareto fronts. In addition, a general pattern can be observed that classical and hybrid designs are generally more appealing and associated with better performance than the correlated designs, and the replicated designs are least desirable among all categories with worst performance for both $E(s^2)$ and $tr(\mathbf{AA'})$ criteria.

CONCLUSIONS

The 16-run screening designs with six to eight factors are popular for cases with tight constraints on design size relative to the number of design factors to be considered. Common choices include classical designs built from a 2^4 factorial design with completely confounded main effects with two- and higher-order interactions, such as the fractional factorial resolution IV design for the six-factor case. Considering that some nonregular designs (Jones and Montgomery 2010) with specially defined design structure can avoid complete confounding between main effects and two-way interactions and allow better evaluation of some of these terms, Johnson and Jones (2010) studied the list of all orthogonal nonisomorphic regular and nonregular designs for six-, seven-, and eight-factor cases and provided catalogs of defining equations for design generation. These catalogs are helpful for knowing what choices exist but do not directly provide guidance about how to choose among these options. Without prior information about which of the two-way or higher-order interactions might be active, it is still challenging to choose from the collection of possibilities based on only the aliasing patterns.

Building on these catalogs of designs, this article uses two common general design criteria, $E(s^2)$ and $tr(\boldsymbol{AA}')$, to compare the candidate choices. These criteria quantify the average degree of confounding and potential bias in the parameter estimates, if some of the two-factor interactions are actually active. A two-stage Pareto front approach is used to facilitate rational decision making. The first objective stage eliminates all inferior options and keeps only nondominated designs on the Pareto front for further consideration. The results from this stage provide a set of candidates from which a best experimental design can be selected. The second subjective stage uses a set of graphical tools to compare the promising designs based on quantitative evaluation of their relative performance and trade-offs as well as robustness to different weighting preferences and provides useful information for making a rational and defensible decision that best matches the priorities of the experiment.

In the subjective stage, in addition to the weighting preference, there are some choices to make for scaling criteria values into a comparable range and combining multiple criteria into a single metric (desirability function) for ranking the design performance. Making these subjective choices will impact the results and hence choices should be made to reflect experimenter's priorities and how much to penalize inferior performance for a given criterion. With the summary of design performance given in the tables, the user could construct alternatives using different scalings and desirability function forms, if so desired. Since the Pareto front remains unchanged regardless of these choices, the Pareto approach offers great flexibility and computational advantage for exploring alternative choices and understanding their impacts, which was demonstrated with the six-factor case.

The results from the three case studies (six-, seven-, and eight-factor situations) indicate that among the four categories of design structures, the replicated designs are least desirable, with the worst performance for both criteria across all cases. In general, the classical and hybrid designs tend to associate with better performance than correlated designs. For each case, only a small proportion of the designs from the population are considered as sensible choices from which to select. However, there is no universal best design for any of the three cases evaluated, and which design to select depends on priorities of the study and how the experimenter prefers to value the two criteria.

The streamlined decision-making process, including how to use the graphical summaries for aiding informed decision making, was illustrated with the three cases for six, seven, and eight factors based on the two general criteria. However, the Pareto front approach for multiple objective optimization can be adapted for applications with more objectives and more complexity in design structure, randomization, and space constraints, as well as metrics for quantifying the objectives. In addition, the Pareto front approach is not limited to evaluating a fixed set of candidate choices and can be generalized to other problems, when it is not practical to exhaustively search and evaluate all possible choices, by using global optimization search algorithms.

ABOUT THE AUTHORS

Dr. Lu Lu is a visiting assistant professor in the Department of Mathematics and Statistics at the University of South Florida. She earned a doctorate in statistics from Iowa State University.

Dr. Mark E. Johnson is a professor of statistics at the University of Central Florida. He is a Fellow of the American Statistical Association, Chartered Statistician with the Royal Statistical Society, and Elected Member of the International Statistical Institute.

Dr. Christine M. Anderson-Cook is a research scientist in the statistical sciences group at Los Alamos National Laboratory in Los Alamos, New Mexico. She earned a doctorate in statistics from the University of Waterloo in Ontario. She is a fellow of both American Society for Quality and the American Statistical Association.

REFERENCES

Booth, K. H. V., Cox, D. R. (1962). Some systematic supersaturated designs. *Technometrics*, 4:489–495.

Bursztyn, D., Steinberg, D. (2006). Comparison of designs for computer experiments. *Journal of Statistical Planning and Inference*, 136:1103–1119.

Cornell, J. (2002). *Experiments with Mixtures: Design, Models, and the Analysis of Mixture Data*, 3rd ed. New York: Wiley.

Derringer, G., Suich, R. (1980). Simultaneous optimization of several response variables. *Journal of Quality Technology*, 12:214–219.

Johnson, M. E., Jones, B. (2010). Classical design structure of orthogonal designs with six to eight factors and sixteen runs. *Quality and Reliability Engineering International*, 27:61–70.

Jones, B., Montgomery, D. C. (2010). Alternatives to resolution IV screening designs in 16 runs. *International Journal of Experimental Design and Process Optimization*, 1(4):285–295.

Lu, L., Anderson-Cook, C. M. (2012). Rethinking the optimal response surface design for a first-order model with two-factor interactions, when protecting against curvature. *Quality Engineering*, 24(3):404–421.

Lu, L., Anderson-Cook, C. M. (2014). Balancing multiple criteria incorporating cost using Pareto front optimization for split-plot designed experiments. *Quality and Reliability Engineering International*, 30:37–55.

Lu, L., Anderson-Cook, C. M., Lin, D. (2013). Optimal designed experiments using a Pareto front search for focused preference of multiple objectives. *Computational Statistics and Data Analysis*, 71: 1178–1192.

Lu, L., Anderson-Cook, C. M., Robinson, T. J. (2011). Optimization of designed experiments based on multiple criteria utilizing Pareto frontier. *Technometrics*, 53:353–365.

Lu, L., Anderson-Cook, C. M., Robinson, T. J. (2012). A case study to demonstrate Pareto frontiers for selecting a best response surface design with simultaneously optimizing multiple criteria. *Applied Stochastic Models in Business and Industry*, 28:206–221.

Lu, L., Chapman, J., Anderson-Cook, C. M. (2013). Selecting a best allocation of new data for improving the estimation precision of system and subsystem reliability using Pareto fronts. *Technometrics*, 55(4):473–487.

Sun, D. X., Li, W., Ye, K. Q. (2002). An algorithm for sequentially constructing non-isomorphic orthogonal designs and its applications. Technical Report SUNYSB-AMS-02–13, Department of Applied Mathematics and Statistics, State University of New York at Stony Brook.