Design Optimization Based on D-optimality for Multiple Responses

Damola Akinlana[1], Lu Lu[2]

[12]Department of Mathematics and Statistics, University of South Florida

Abstract

Due to various resource restrictions, there has been increased needs to measure multiple responses in a single experiment. Central composite designs have been broadly used for estimating response surfaces from experiments with multiple responses, which typically require a large number of runs across the super space of all relevant design factors. However, in many of these experiments, each individual response is often affected by only a subset of design factors, and such information might be obtained from earlier screening experiments. We propose a more cost-efficient design selection strategy based on utilizing this prior knowledge and the Pareto front approach to select D-optimal designs with balanced performance on multiple responses. A Pareto aggregate coordinate exchange algorithm has been adapted to efficiently identify the Pareto front based on D-efficiencies measured for multiple responses. The method is illustrated with two examples and compared with existing methods on a variety of design characteristics.

Keywords: Multiple Responses, Optimal Designs, Coordinate Exchange Algorithm, Pareto Front, D-optimality

1. Introduction

In modern design of experiments, there has been an increased demand for measuring multiple responses in a single experiment. The goal in experimental design with multiple responses is to select appropriate designs that perform well for all the responses. Quite a number of design strategies have been proposed in the literature for designed experiments with one response variable and one or more design factors. Existing designs for multiple responses often are built in the super space of all design factors that are relevant to multiple responses and hence generally require more experimental runs than what are necessary for efficiently estimating each response. There are limited works for the case of multiple responses where different subsets of design factors might be needed to model each response.

Response surface models are common in settings where characteristics of the system quantified by the response variables are to be optimized (Myers et al, 2016). One of the most common classical designs for estimating full second order polynomial regression multiple responses models is the Central Composite Designs (CCDs) (Box and Wilson, 1951). This design is a combination of three parts: a two-level full or fractional factorial points, axial points and center points. The resolution of a fractional factorial design (Box and Hunter, 1961) describe how the effects in a fractional factorial design are aliased with other effects. Resolution III, IV, and V designs are the common resolutions in literature used for fractional factorial designs. For resolution III designs, main effects are aliased with second order effects. For resolution IV, no main effects are aliased with second order effects, but some two-factor interactions are aliased with each other. Resolution V fractional factorials are commonly used in CCDs because they are the smallest fractional factorials that allow estimation of all first and second order effects. Because designs of greater resolutions require more experimental runs, fractional factorials of resolutions higher than V are usually avoided in CCDs. The axial points are experimental runs where all factors are held constant at zero except for one factor, which is set to be $\pm\alpha$, where α is a design parameter. There are $2k$ axial points, along with the center runs (experimental runs where all factors are set to zero) which allow for estimation of model parameters in squared terms.

The CCD described above can be used in estimating full quadratic response surface models where all factors are treated symmetrically. However, for design problems with multiple responses, screening designs could be employed to identify active design factors for each response. For instance, with the information from screening experiment, some design factors may not be expected to have much impact on a certain response variable, or some two-factor interactions are unlikely to be active for a particular response model. Marget and Morris (2019) proposed the Unique Factor Central Composite Designs (UF-CCDs), which utilizes the information from the screening experiment to guide a more efficient response surface design. They consider situations where screening experiments or process knowledge suggests that only a subset of the factors is needed in modelling each response, and where these subsets are generally different for each response. The modification made to the CCD reflects in how the fractional factorial portion of the design is selected. As explained earlier, a full factorial in all factors is often a much larger design than what is needed for estimation of each response model. Given r response variables and k design factors, the fractional portion of the UF-CCDs were constructed such that for any given response, the factors associated with that response form either a full factorial or a resolution IV or resolution V fractional factorial. That is, the design will be a full or fractional factorial

for the factors related to each response, but not for all k factors altogether, and so will require fewer runs than the traditional CCDs. The UF-CCDs also take advantage of intentional aliasing of some of the main effects. Since not all factors appear in each response model, pairs of factors that do not appear in any model can be aliased, permitting further reduction in the size of the design. More on the discussion of UF-CCDs is available in Marget and Morris (2019).

The CCDs and UF-CCDs are both classical designs. Although classical designs are mostly used in the industry, they are generally less flexible in terms of the number of required design size. Computer optimal design methodology was developed to generate designs when user constraints preclude the use of classical designs. Computer-aided designs are experimental designs that are generated based on an optimality criterion and are generally referred to as optimal designs (Jin et al, 2003). As an alternative to CCDs and UF-CCDs, this paper considers design optimization based on D-optimality for multiple responses by utilizing the Pareto frontier approach for multiple objective optimization. Particularly, we aim to simultaneously optimize D-optimality for multiple responses surface models which involve different subsets of design factors by seeking the Pareto front of D-efficiencies for multiple responses. In order to efficiently identify the Pareto front, this article utilizes a customized Pareto Aggregate Coordinate Exchange (PACE) algorithm which was adapted from the PAPE algorithm from Lu et al. (2011) and Cao et al. (2017).

The rest of the article is organized as follows. Section 2 discusses D-optimality criteria use for optimizing multiple responses. Section 3 describes the Pareto aggregate coordinate exchange optimization algorithm for generating the Pareto front of multiple criteria. Section 4 presents examples of using the PACE algorithm for selecting optimal designs with different numbers of design factors and demonstrates advantages of multiple responses D-optimal designs compared to existing methods. Section 5 renders the concluding remarks.

2. D-optimal Designs with Multiple Responses

In generating optimal designs where the focus is on precise parameter estimations, D-optimality is the most popular design optimality criterion due to simplicity of its implementation and computation. D-optimal designs are designs that maximize the determinant of the design moment, which is equivalent to minimizing the volume of the confidence region of the regression parameters (Meyers et al, 2016). The D-criterion is defined as:

$$|M(\boldsymbol{\xi})| = |X'(\boldsymbol{\xi})X(\boldsymbol{\xi})|/N^{k+1} \qquad (1)$$

where $X(\boldsymbol{\xi})$ is the $N \times (k+1)$ model matrix for design $\boldsymbol{\xi}$ with k design factors and $\boldsymbol{\beta}$ is the associated $(k+1) \times 1$ vector of regression coefficients. The D-optimal design maximizes $|M(\boldsymbol{\xi})|$ over the design space spanned by all possible designs, $\boldsymbol{\xi \epsilon \Omega}$. The D-efficiency of a design, $\boldsymbol{\xi}$ is given by:

$$D_{eff}(\boldsymbol{\xi}) = (|M(\boldsymbol{\xi})|/|M(\boldsymbol{\xi}_{\boldsymbol{D}}^{*})|)^{1/k+1} \qquad (2)$$

where $\boldsymbol{\xi}_{\boldsymbol{D}}^{*}$ is the D-optimal design that maximizes $|M(\boldsymbol{\xi})|$. D-optimal designs are often generated by an iterative search algorithm and seeks to minimize the covariance of the parameter estimates for a specified model. It is worth mentioning that other commonly used optimality criterions in the literatures include; the A-criterion, G-criterion and I-criterion (Rady et al, 2009). These criterions seek to minimize the average variance of parameter estimates or the prediction variance across the design space. Specifically, the A-optimality criterion is defined as $min\{trace(X'X)^{-1}\}$ and it is equivalent to minimizing

the average variance of the estimated regression coefficients. The A-efficiency of a design $\boldsymbol{\xi}$ is defined as; $A_{eff}(\boldsymbol{\xi}) = tr[M^{-1}(\boldsymbol{\xi}_{\boldsymbol{A}}^{*})]/tr[M^{-1}(\boldsymbol{\xi})]$, where $\boldsymbol{\xi}_{\boldsymbol{A}}^{*}$ is the A-optimal design. The G-optimality criterion is defined as:

$$\min_{x_i, i=1,\ldots,n} \max_{x \in \chi} var(\hat{\boldsymbol{y}}_{\boldsymbol{x}}),$$

where

$$var(\hat{\boldsymbol{y}}_{\boldsymbol{x}}) = \sigma^2 f^T(x)(X^T X)^{-1} f(x).$$

Hence, the G-optimal design seeks to minimize the maximum prediction variance over the design space. The G-efficiency of a design $\boldsymbol{\xi}$ is defined as; $G_{eff}\ (\boldsymbol{\xi}) = p \ / \max_{x \in \chi} var(\hat{\boldsymbol{y}}_{\boldsymbol{x}})$, where p is the number of parameters in the model. The criterion that minimizes the average prediction variance across the design space is regarded as the I-optimality criterion and it can be define as $trace[(X'X)^{-1}B]$, where $B = \int_{\chi} f^T(x) f(x) dx$ is the moment matrix of the region of interest. That is, I-optimality criterion minimizes the average prediction variance by integrating over the design space. The I-efficiency of a design $\boldsymbol{\xi}$ is defined as; $I_{eff}(\boldsymbol{\xi}) = I(\boldsymbol{\xi}_{\boldsymbol{I}}^{*})/I(\boldsymbol{\xi})$, where $\boldsymbol{\xi}_{\boldsymbol{I}}^{*}$ is the I-optimal design. Note that, all the D-efficiency, A-efficiency, G-efficiency, and I-efficiency criteria, measure the design performance relative to the optimal choice based on each optimality criteria, and hence take values within the [0,1] range.

From earlier works on finding optimal designs for design problems with multiple responses, some optimal design strategies have been proposed. Fedorov (1972) proposed an extension of *D*-optimality for multiple responses called *MD*-optimality which requires that the variance-covariance matrix of the response be known, this assumption is often not realistic in practice. Cooray-Wijesinha and Khuri (1987) suggested an alternative where the variance-covariance matrix is estimated with an initial design and then used to add points to the design. Although their approach avoids Fedorov's assumption, it may pose difficulty in implementation if sequential operation is inconvenient. Chang (1997) suggested a design that does not share either of these shortcomings but does not offer the flexibility to use different models for each response as suggested by screening experiments or previous knowledge. His proposed design allows for each response to have either a complete first or second order model, but all response models contain all factors.

As an improvement to these earlier works on design problems with multiple responses, the objective in this work is to generate D-optimal designs such that the D-efficiency values of multiple responses are maximized simultaneously. This can be viewed as a multiple objective optimization problem where the goal is to simultaneously optimize the D-efficiency for each response. In the derivation of the D-criterion for each response model, $X(\boldsymbol{\xi})$ from equation (1) describe the design matrix for each response model. As described in examples in section 4, prior knowledge from an earlier screening design about which factors affecting which response is used to determine the design matrix for each response model. Thus, for a particular response model, the first order terms, two factor interaction terms and squared terms of the influencing factors form the model matrix for that response. Therefore, the design matrix for each response varies due to the difference in the subsets of factors influencing each response and potential active effects.

Various techniques for obtaining the optimal set for multiple criteria design problem have been proposed in the literature. In the works of DuMouchel and Jones (1994) and Allen et al (2003), criteria based on mean squared error were developed in an effort to

simultaneously minimize variance and bias. Jones and Nachtsheim (2011) used conditional optimization to find designs which optimize one objective while achieving at least some threshold for a secondary criterion. The desirability function approach (see Harrington 1965 and Derringer and Suich 1980) considers optimizing a linear combination of two or more criteria. Linear combinations are formulated by predefined weighting schemes where the weights reflect the relative importance of each of the criteria. Optimization algorithms such as direct search methods (Lewis et al, 2000), mathematical programming approaches (Del Castillo and Montgomery 1993), simulated annealing (Kirkpatrick et al, 1983), or the genetic algorithm (Holland 1975) are popular choices when desirability functions are utilized. Although the desirability function approach has been used extensively in literature, it fails to consider the trade-offs between criteria directly and its results depend heavily on the user selected weights and weighting scheme. Experimenting with different weight choices can be time consuming and computationally intensive since every set of weights requires a separate search for an optimal solution.

An alternative to the desirability approach in finding the optimal set for multiple criteria design problem is the Pareto front optimization approach. Pareto optimization has been extensively used for optimizing multiple responses in many disciplines (Gronwald et al, 2008; Trautmann and Mehnen, 2009). In design of experiment, the Pareto front approach simultaneously considers multiple objectives by constructing a frontier of competitive designs while explicitly considering tradeoffs between opposing criteria. Various algorithms have been proposed for populating the Pareto fronts of designs and for evaluation and comparison of designs identified on the front (Park, 2009; Lu et al, 2011; Sambo et al, 2014). A major advantage of this approach is that, the search for solutions occurs only once and then different metrics can be easily explored to select optimal designs for various scenarios. In this work, we adopt the Pareto front optimization approach for design of experiments with multiple responses.

Without loss of generality, the general goal of a multiple criteria design optimization problem is to maximize C (≥ 2) criteria simultaneously given constraints on the input factors. Let $\boldsymbol{\xi} = (\boldsymbol{d}_1', \boldsymbol{d}_2', \dots, \boldsymbol{d}_N')' \epsilon \boldsymbol{\Omega}$ denote a design matrix of dimension $N \times k$ where N is the number of design points and k is the number of design factors; the set of all possible $N \times k$ design matrices for a given candidate set of points is denoted by $\boldsymbol{\Omega}$. The candidate set is a collection of treatment combinations from which the search algorithm chooses the treatment combinations to include in the design. Let $y = F(\boldsymbol{\xi}) = \left(f_1(\boldsymbol{\xi}), f_2(\boldsymbol{\xi}), \dots, f_C(\boldsymbol{\xi})\right)^T$ denote the vector of criteria values corresponding to the design matrix, $\boldsymbol{\xi}$. Then, the space containing all obtainable criteria vectors is called the criterion space. A solution $\boldsymbol{\xi}_1$ is said to *Pareto dominate* another solution $\boldsymbol{\xi}_2$ if $f_j(\boldsymbol{\xi}_1) \geq f_j(\boldsymbol{\xi}_2)$ for all $j \in \{1,2, \dots, C\}$ and there exists at least one $j \in \{1,2, \dots, C\}$ such that $f_j(\boldsymbol{\xi}_1) > f_j(\boldsymbol{\xi}_2)$. In this case, the criteria vector $F(\boldsymbol{\xi}_2)$ is said to be *dominated* by $F(\boldsymbol{\xi}_1)$. In this work, the criteria vector corresponding to a particular solution is referred to as a *point* in the criterion space. A solution is *Pareto optimal* if and only if no other solution dominates it and its corresponding criteria vector is a *non-dominated* vector. We refereed to the *Pareto optimal set* as the set of Pareto optimal solutions and the corresponding set of criteria vectors as the *Pareto front*. Marler and Arora (2004) provides a good overview of the Pareto front concepts.

3. Pareto Aggregate Coordinate Exchange (PACE) Algorithm

Enumerating all possible designs in an experimental region tends to pose a challenge even for a region of moderate size. As an alternative to this, an initial design can be considered and successfully improved upon by exchanging existing design points with candidate points which produce improvement in the design criteria. The exchange algorithms are the most used algorithms for optimizing design criteria. Although proving to be highly effective, applications of the exchange algorithms can be computationally prohibitive for very large problems and in situations for which design criteria are sufficiently complex; thereby precluding efficient evaluation (Rady et al, 2009). There are generally two types of exchange algorithms; the point exchange and the coordinate exchange algorithms (Nguyen and Miller, 1992). Although point exchange and coordinate exchange algorithms are deterministic heuristics search methods, their implementation in practice generally involves a random component where multiple random starts are often used in order to increase the probability that the global optimum is attained. Rows in the design matrix are exchanged in the point exchange algorithm, while coordinates of the design matrix are exchanged in the coordinate exchange algorithm. Another difference is that the point exchange procedure requires a candidate set while the coordinate exchange procedure does not.

The point exchange procedure searches the entire candidate set to replace each row in the design matrix by another location from the candidate set to improve at least one of the criteria without deteriorating the others. This search procedure can become computationally intensive and even prohibitive as the number of design factors increases, especially for factorial design spaces that could grow exponentially in the number of factors. Because the point exchange only updates the current design with a strict improvement on the criteria values, designs which do not dominate and are not dominated by the current design are discarded, when in fact these might be potential solutions in the Pareto optimal set and could be optimal solutions when different user priorities are selected. Although repeating the search with multiple starts might find more points, however finding all points on the front can require a very large number of starts and potentially repeated reevaluation of same designs across the search with multiple random starts. Lu et al (2011) proposed a modification to the point exchange algorithm, described as Pareto Aggregate Point Exchange (PAPE) to improve the efficiency of the regular point exchange by keeping track of all the non-dominated points and building the Pareto front along the searching process. Thus, the PAPE algorithm efficiently explores the design space by populating the Pareto frontier with all possible contending designs identified during the search.

The coordinate exchange on the other hand does not search the entire candidate set for a replacement. That is, it does not require a candidate set, but instead it randomly generates a starting design and exchanges each entries of the design matrix to search for an improvement. The absence of a candidate set for the coordinate exchange procedure reduces demands on computer memory which makes the coordinate exchange computationally more efficient than the point exchange. For design problems with more than a few factors, or in applications with large candidate sets, the PAPE can be computationally inefficient if not prohibitive (Cao et al, 2017). Due to the complexity of the factorial design spaces of the examples considered in this work and to avoid the limitations of the PAPE algorithm, the Pareto frontier approach developed in this article uses a coordinate exchange algorithm (hereafter refer to as PACE).

The PACE algorithm utilizes the idea of the PAPE algorithm and adapts it to searching for the Pareto front using the coordinate exchange. Cao et al (2017) also proposed a Pareto based coordinate exchange operator, which is different from the PACE with a different second stage comparison. The PACE algorithm for finding the Pareto front with a single random start begins with a randomly generated initial design with a nonsingular moment matrix and then replaces each coordinate in the design matrix by another value in the levels of design factors considered to improve at least one of the criteria without deteriorating the others. In this work, the design space is the grid of points generated from the levels of design factors. For instance, consider a design with five factors and each factor has five levels, then the design space is a full five-factors five-levels factorial design points. Each random start is randomly selected from this grid of points. An overview of the PACE algorithm used in this work is described by the following steps:

(1) From a given candidate set, randomly generate an initial design of N runs and k factors with a nonsingular moment matrix ($|X'X| \neq 0$). For this current design, $\boldsymbol{\xi}$, evaluate the user-specified C-dimensional criterion vector, $F(\boldsymbol{\xi}) = \left(f_1(\boldsymbol{\xi}), f_2(\boldsymbol{\xi}), \dots, f_C(\boldsymbol{\xi})\right)^T$.

(2) Initialize two null sets: the set of Pareto designs denoted as P and the set containing the corresponding Pareto fronts, denoted as PF ; then add $\boldsymbol{\xi}$ to P and add $F(\boldsymbol{\xi})$ to PF.

(3) For the current design $\boldsymbol{\xi}$, the PACE algorithm is carried out as follows; for $i = 1$ to N and $j = 1\ to\ k$, swap each (i, j) coordinate of $\boldsymbol{\xi}$ with the remaining levels of factor j to produce a new design $\boldsymbol{\xi}^*$. Two sets of comparisons are then made:

(i) First comparison: updating the current design.
This first comparison is between the current design $\boldsymbol{\xi}$ and the new design $\boldsymbol{\xi}^*$ to determine if the current should be replaced by the new one. If the new design improves at least one of the criteria without deteriorating any other criteria (i.e if $\boldsymbol{\xi}^* > \boldsymbol{\xi}$), then the current design is replaced with the new one (i.e., $\boldsymbol{\xi} = \boldsymbol{\xi}^*$). This procedure is done for every coordinate in the design until no improvements can be made.

(ii) Second comparison: updating the Pareto front and the Pareto optimal set.
The second comparison is between the new design, $\boldsymbol{\xi}^*$ and the ones in the "current" set of non-dominated designs P. If $\boldsymbol{\xi}^*$dominates at least one of the designs in the "current", then add $\boldsymbol{\xi}^*$ to P and remove the designs dominated by $\boldsymbol{\xi}^*$. If $\boldsymbol{\xi}^*$neither dominates nor is dominated by any designs in the current generation of P, then just add $\boldsymbol{\xi}^*$to P. If $\boldsymbol{\xi}^*$ is dominated by at least one of the designs in the "current" set, then discard $\boldsymbol{\xi}^*$ and no update is needed for the current Pareto set.

The procedure above is stopped when the current designs cannot be improved upon. At the end of the search, there will be m non-dominated designs in the set $P(i.e\ P = \{\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \dots, \boldsymbol{\xi}_m\})$ and m associated criterion vectors in the set PF $(i.e\ PF = \{F(\boldsymbol{\xi}_1), F(\boldsymbol{\xi}_2), \dots, F(\boldsymbol{\xi}_m)\})$.

(4) Repeat steps 1 to 3 with S different random starts (S is usually chosen to be a large number to ensure the identification of a complete Pareto front). The set of non-dominated designs are updated by combining the Pareto fronts obtained from different random starts.

4. Examples

In this section, we illustrate the proposed method and the PACE algorithm using two examples which involve five and ten design factors, respectively, and multiple responses. We then compare the selected multiple responses D-optimal designs with the UF-CCDs developed by Marget and Morris (2019).

4.1 Example 1: Five Factors with Four Responses

Consider the results of a screening experiment summarized in Table 1. This is the same screening experiment which was discussed in Marget and Morris (2019) (see Table 3 of the paper)). A screening experiment is a study focused on identifying the factors that influence each response, rather than a more detailed study to quantify those relationships. An X in the table indicates that the response associated with that row is related to the factor associated with that column.

Table 1: Five factors screening experiment

	Factor				
Response	1	2	3	4	5
1	X	X	X		
2		X	X	X	
3	X		X		X
4	X			X	

The goal here is to select a 20-run D-optimal design that simultaneously perform well for estimating all four response variables. We consider five levels for each design factor at -2 , -1,0,1,2, in order to make a direct comparison with the Resolution IV Unique Factor CC D (Res IV UF-CCD) from Marget and Morris (2019). This has resulted in a candidate set of $5^5 = 3125$ design points. The starting design was randomly selected from the candida te set. By using 20000 random starts, the PACE algorithm resulted in a single dominating point on the Pareto front. No other design has been found to have higher D-efficiency val ues for any of the four responses after running up to 40000 random starts. Our interest her e is in precise estimation of the parameter estimates of each response model. Thus, for a g eneral linear regression model given by;

$$Y = X\beta + \varepsilon,$$

where, $var(\varepsilon) = \sigma^2 I$, the standard deviation of the estimate of any element of β is σ times the square root of the corresponding diagonal element of $(X^T X)^{-1}$. If the estimated standard deviation of a parameter of a model is denoted by s and the unknown true standard deviation is σ, then the normalized standard deviation is define as $\frac{s}{\sigma}$, and it is only a function of the design matrix, X. This quantity reflects the performance of inference that can be expected from a given design. For each response, we fit a second order linear regression model and then estimate the standard deviations of the model coefficients. Figure 1 shows the averages of normalized standard deviations for all term coefficients (all effects), first-order term coefficients (main effects), interaction term coefficients (interaction effects), and squared term coefficients (squared effects) for the D-optimal design obtained and that of the Res IV UF-CCD. These indices are estimated to reflect how well a given design performs with respect to estimation of different groups of coefficients in each response model.

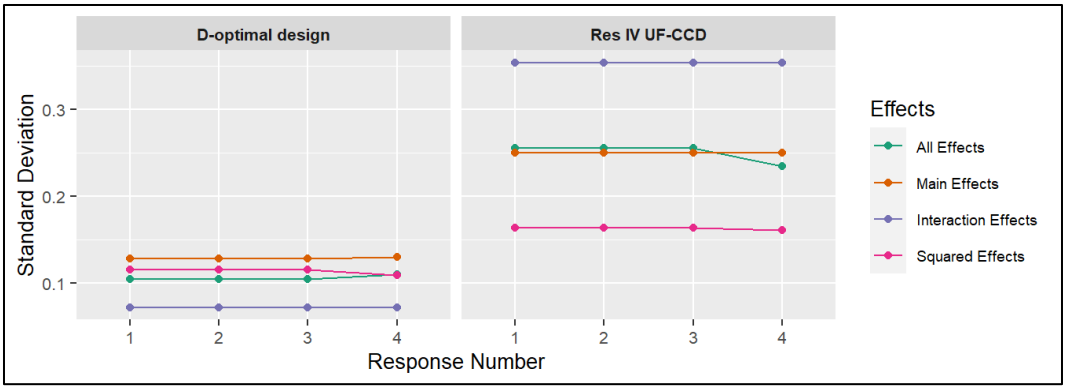

Figure 1: Averages of standard deviations of parameter estimates for the five factors example.

As observed in Figure 1, the standard deviations of model coefficients for each response model are substantially higher for the Res IV UF-CCD than that of the D-optimal design. This implies that the structure of the D-optimal design is more efficient in terms of precise parameter estimation than the Res IV UF-CCD. To further compare the performance of the designs, we obtained response-based D-efficiency, A-efficiency, G-efficiency and I-efficiency values for the D-optimal design and the Res IV UF-CCD. In Figure 2, it can be seen that, the efficiency values of these criteria for each response model are significantly higher for the D-optimal design, implying that the D-optimal design generally outperforms the Res IV UF-CCD. For both designs, the D-efficiency value for response 4 is slightly lower than those of the remaining responses and its A, G and I efficiency values are considerably higher than those of the others. We suspect that this may be as a result of reduced number of parameters for response 4 model since from the screening experiment, only two factors were identified to influence response 4 while three factors were expected to influence the remaining three responses.

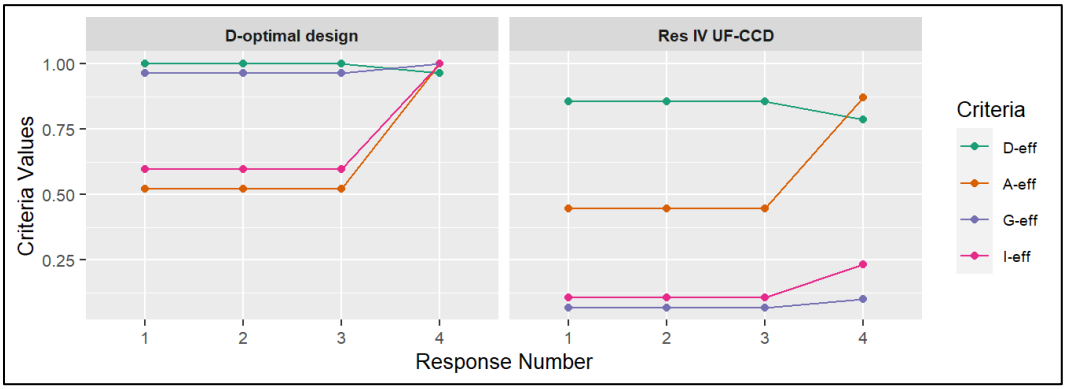

Figure 2: The plot showing the D-eff, A-eff, G-eff and I-eff for the five factors D-optimal design and Res IV UF-CCD

The geometric structure of the D-optimal design obtained for the five factors screening experiment with multiple responses is compared with that of the Res IV UF-CCD (see appendices A and B). The structure shows that there is no center run selected for the D-optimal design and most of the runs are located at the edge of the design space. This is expected because in order to optimize D-efficiencies of multiple responses, more design

points are likely to be pushed to the edge rather than located around the center of the design space. Particularly, it can be observed that all the runs of the D-optimal design are combinations of the levels; (-2,0,2) while for the Res IV UF-CCD, some of the runs are combinations of the levels (-1,1) and some are combinations of the levels (-2,0,2).

4.2 Ten Factors Design with Four Responses

This section considers another example from Marget and Morris (2019) with ten design factors and four response variables. The result from the screening experiment is summarized in Table 2 (a recap of Table 2 from Marget and Morris (2019)).

Table 2: Ten factors screening experiment

	Factor									
Response	1	2	3	4	5	6	7	8	9	10
1	X	X	X	X	X					
2			X		X	X	X			
3			X	X			X	X		
4				X				X	X	X

The interest here is to select a 30-run D-optimal design that offers balanced performance o n all four response variables. We again consider five levels for each design factor and hen ce the candidate set would include $5^{10} = 9765625$ design points, which will prohibit the u se of PAPE algorithm with this large size of the candidate set. A Resolution V Unique Fa ctor CCD (Res V UF-CCD) was proposed by Marget and Morris (2019), whose design m atrix is shown in Appendix C. Unlike the five factors case, the PACE algorithm for this te n factors case did not find a single dominating design on the Pareto front, instead we obtai ned a suite of competing non-dominated designs. By running up to 100000 random starts, the PACE algorithm identified a Pareto front consisting of eight competing designs. To fu rther examine the characteristics of the eight designs identified on the Pareto front, Figure 3 shows the average normalized standard deviations of model coefficients of each respons e for the eight D-optimal designs and Res V UF-CCD (denoted as design number 9). As w ith the five factors case, the result indicates that the D-optimal designs have significantly l ower standard deviations, implying that the D-optimal designs obtained are generally mor e efficient for producing precise estimates of model parameters.

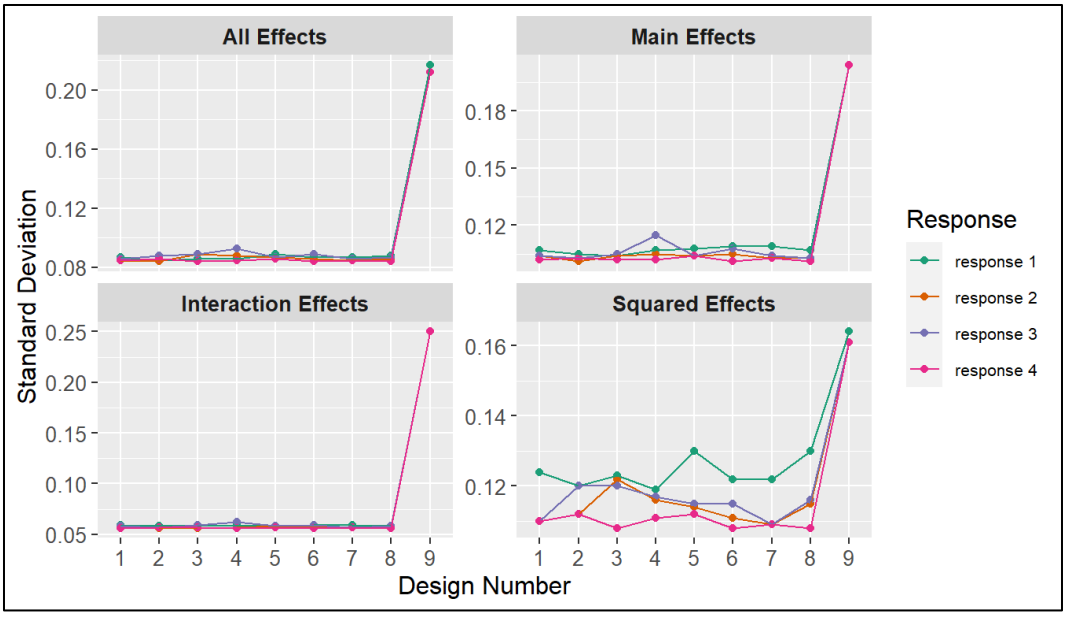

Figure 3: Averages of standard deviations of parameter estimates for the ten factors example with 30 experimental runs

Although the competing designs on the Pareto fronts clearly outperform the UF-CCD designs based on the estimated standard deviations of parameter estimates, however; for the purpose of decision making, it is important to consider the trade-offs between some design objective criteria for the competing designs. Moreover, it is helpful to consider additional secondary criteria on other design characteristics to ensure a good overall performance of the final selected designs. Consider the scenario in which a researcher wishes to select a designed experiment for studying the relationship between the four responses and the ten potential factors $(X_1 - X_{10})$. In other to minimize cost of experiment, screening experiment was conducted to obtain information about factors with the most significant impacts on each response as given in Table 2. As a way of illustration, the specified model with all main effects, interactions effects and squared effects for response 1 can be written as:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_{12} X_{12} + \beta_{13} X_{13} + \beta_{14} X_{14} + \beta_{15} X_{15} + \beta_{23} X_{23} + \beta_{24} X_{24} + \beta_{25} X_{25} + \beta_{34} X_{34} + \beta_{35} X_{35} + \beta_{45} X_{45} + \beta_{11} X_{11} + \beta_{22} X_{22} + \beta_{33} X_{33} + \beta_{44} X_{44} + \beta_{55} X_{55} \quad (3)$$

The specified model for the remaining responses can be obtained in a similar way. Suppose the goal of the experiment is to obtain optimal designs that estimate each response as precisely as possible, and we also want to protect against the impact of potential misspecification on the estimation of coefficients and error variance if any of the omitted factor is potentially active. Let the specified model for a particular response be $y = X_s\beta_s + \varepsilon$ where X_s is an $N \times k_s$ design matrix of factors specified for that response (based on the result of the screening experiment) while the full response surface model for all design factors is $y = X_s\beta_s + X_o\beta_o + \varepsilon$. In our application, the set of terms $X_s\beta_s$ contains the first order terms, two factor interaction terms and squared terms of factors specified for a particular response while $X_o\beta_o$ includes the first order terms, two factor interaction terms and squared terms of factors not specified for the response. Note that X_o is an $N \times k_o$ design matrix of factors omitted for a particular response.

In the estimation of each response model, the bias vector of the specified model parameter estimates is given by

$$E(\hat{\beta}_s) - \beta_s = [\beta_s + (X_s^T X_s)^{-1} X_s^T X_o \beta_o] - \beta_s = A\beta_o \quad (4)$$

where $A = (X_s^T X_s)^{-1} X_s^T X_o$ is the alias matrix (Myers et al, 2016). A measure of transmitted bias to the estimated coefficients is the sum of the squared transmitted bias (SSB) given by $SSB = \beta_o^T A^T A \beta_o$, where β_o is unknown and often assumed to follow the multivariate normal distribution, $\beta_o \sim N(0, \sigma_{\beta_o}^2 I)$ (Draper and Guttman 1992). An optimal design for minimizing the impact of model misspecification on the estimated coefficients is the one that minimizes $tr(AA^T)$ (Bursztyn and Steinberg, 2006). The bias on the estimated coefficients minimization efficiency for a particular design is given by $tr(A_{opt} A_{opt}^T)/tr(AA^T)$, where 'opt' designation denotes that the matrix represent optimal design according to the criteria. Bias can also be transmitted to the estimate of σ^2 due to model misspecification since;

$$E(MSE_{user}) - \sigma^2 = \beta_o^T [X_s A - X_o]^T [X_s A - X_o] \beta_o / k_s$$

$$= \beta_o^T R^T R \beta_o / k_s \quad (5)$$

where MSE_{user} denotes the residual mean squared error from the misspecified model and $R = X_s A - X_o$. An optimal design for minimizing the impact of misspecification on the estimated error variance is the one that minimizes $tr(R^T R)$(Myers et al, 2016). The bias on the estimated error variance minimization efficiency for a particular design is given by $tr(R_{opt}^T R_{opt})/tr(R^T R)$.

For the competing designs in the Pareto front , the D-efficiency values for overall precision of model parameter estimates; $tr(AA^T)$-efficiency values for examining the impact of model misspecification on the estimated coefficients; and $tr(R^T R)$-efficiency values for examining the impact of model misspecification on the estimated errors were estimated for each response model. Similar to the result of the five factors case, the result in Figure 4 shows that, the eight optimal designs found on the Pareto front for the ten factors example have higher D-efficiency values than the Res V UF-CCD design, thus satisfying our goal of precise parameter estimation. From the result, we can say design 6 performed better for response 1 and design 5 performed better for response 2, response 3 and response 4 respectively when considering the trade-off between the criteria values. Overall, design 5 appears to perform best in terms of offering the best protection against model misspecification for all the four responses.

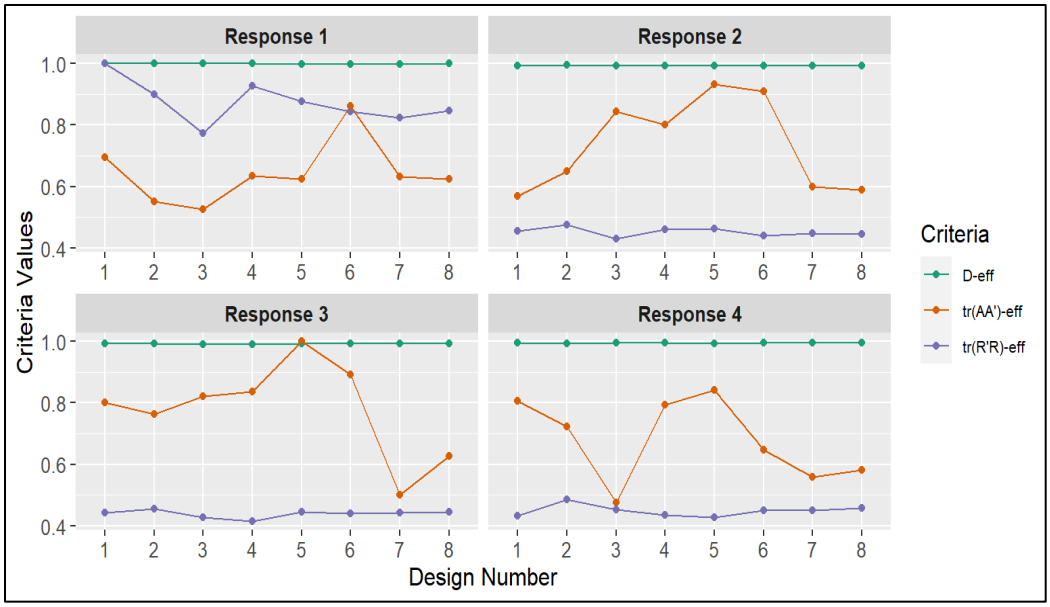

Figure 4: The plot showing the efficiency of D, $tr(AA^T)$ and $tr(R^TR)$ for the ten factors example with 30 experimental runs

If the interest is instead on the prediction ability of the fitted models, then we can examine the G- and I-efficiencies which aim to minimize the maximum prediction variance and the average prediction variance, respectively. We observed that designs 5 and 7, design 8, design 4 and design 1, appears to perform better for response 1, response 2, response 3 and response 4 respectively as shown in Figure 5. We also explore a case with 46 runs. By using the PACE algorithm, twelve competing D-optimal designs were found on the Pareto front. Due to the limitation of space, we omit the detailed results. However, the general patterns are consistent with what we saw for the 30-runs design example.

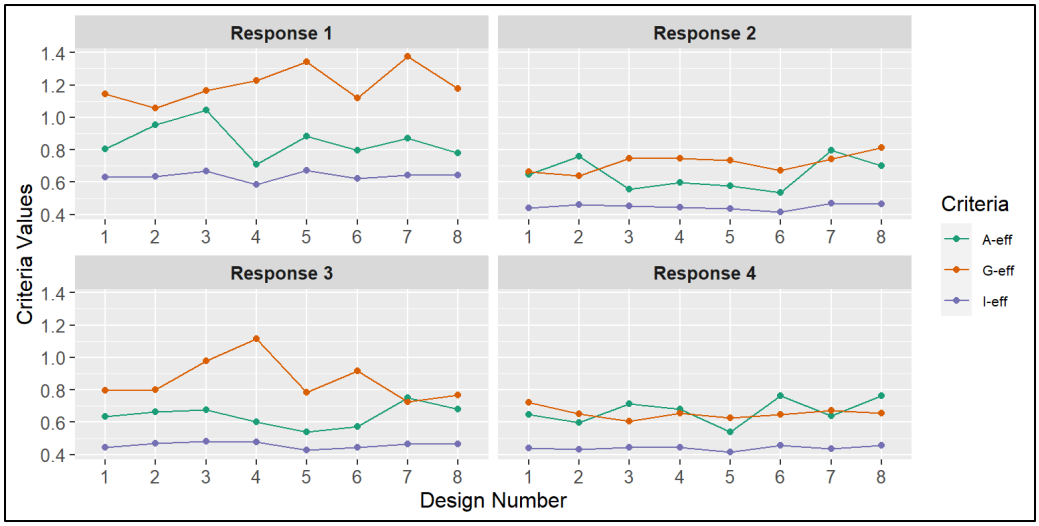

Figure 5: The plot showing A, G and I efficiencies for the ten factors example with 30 experimental runs

Finally, worth mentioning is the computational efficiency of the PACE algorithm used. The algorithm takes an average of forty-eight seconds to run 100 starts on a standard

desktop computer for the five factors examples and five minutes for the ten factors example. Multiple random starts can be run in parallel to improve the computational efficiency. For the five factors example, the Pareto front was identified after running 20000 random starts. Further increasing the number of random effects up to 40000 did not result in any change in the identified Pareto front. For the ten factors example, the number of points found on the pareto front initially increases as the number of random starts increases and then it begins to stabilize after 20000 random starts upward.

5. Concluding Remark

It is not unusual that in many industrial and engineering experiments which involve multiple responses, there is prior information on which subset of the factors is more likely to relate to which response. This prior information could be from the result of a screening experiment or from subject matter expert knowledge. Leveraging this information can lead to improved design selection and thus facilitate improved decision making. The Pareto front optimization can be applied as an alternative to the Central Composite Design approach for experimental design problems with multiple responses. This method allows us to seek optimal designs that simultaneously optimize the precision of estimated response surface models involving different design factors.

This paper examined design optimization based on D-optimality for multiple responses. Specifically, we adapted the PACE algorithm for seeking D-optimal designs with multiple responses and then compare the results with existing methods such as the recently developed UF-CCDs. Both examples with different input spaces demonstrated that the selected D-optimal designs perform better than the UF-CCDs in terms of precise parameter estimation of model parameters. When multiple choices are presented on the Pareto front, further selection of the final design based on evaluating secondary criteria is illustrated through the examples.

Appendix

A.

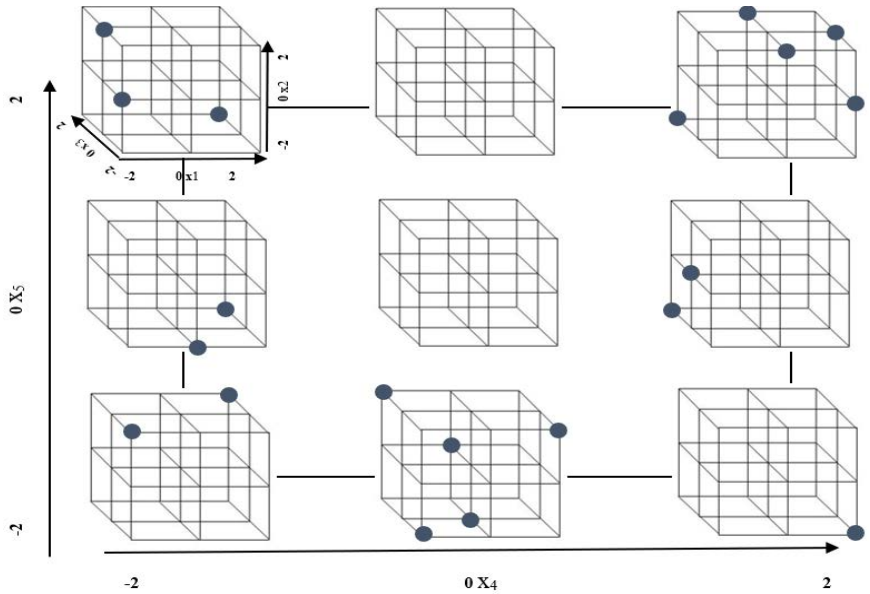

Figure 6: Geometric representation of the five factors D-optimal design

B.

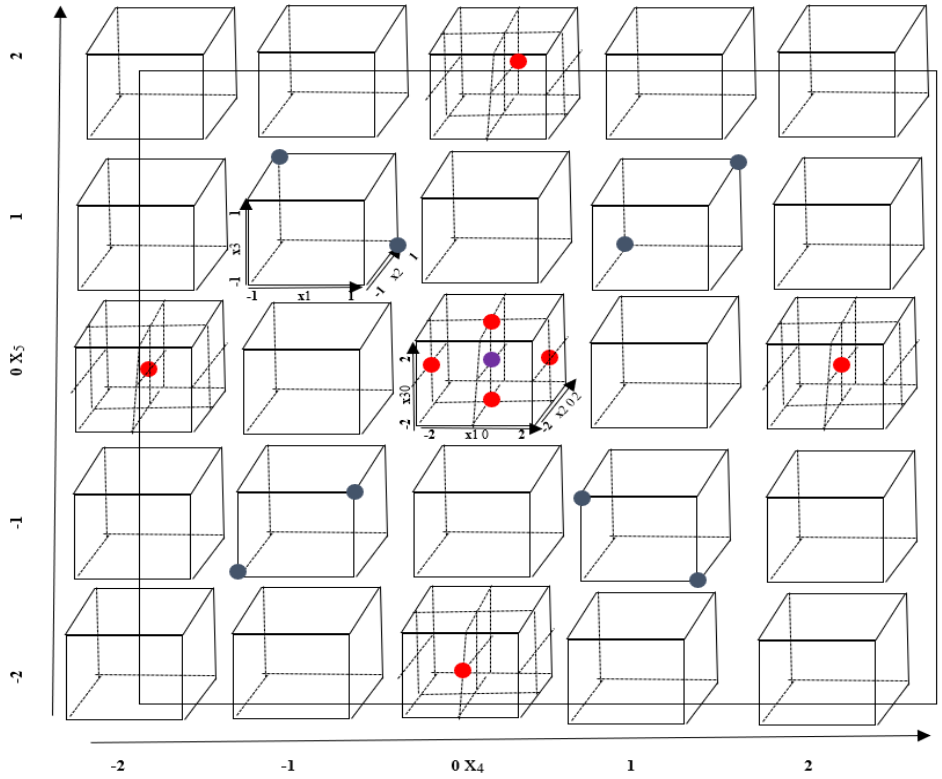

Figure7: Geometric representation of the Res IV UF-CCD. The blue, red and purple circles signifies the corner points, axial points and center points respectively

C.

Table 3: Design Matrix for the Res V UF-CCD design

[1,]	-1	-1	-1	-1	1	-1	-1	-1	-1	-1
[2,]	1	-1	-1	-1	-1	1	1	1	1	-1
[3,]	-1	1	-1	-1	-1	-1	1	-1	1	-1
[4,]	1	1	-1	-1	1	1	-1	1	-1	-1
[5,]	-1	-1	1	-1	-1	-1	-1	-1	-1	1
[6,]	1	-1	1	-1	1	1	1	1	1	1
[7,]	-1	1	1	-1	1	-1	1	-1	1	1
[8,]	1	1	1	-1	-1	1	0	1	0	1
[9,]	-1	-1	-1	1	-1	-1	0	-1	0	-1
[10,]	1	-1	-1	1	1	1	2	1	2	-1
[11,]	-1	1	-1	1	1	-1	-2	-1	-2	-1
[12,]	1	1	-1	1	-1	1	0	1	0	-1
[13,]	-1	-1	1	1	1	-1	0	-1	0	1
[14,]	1	-1	1	1	-1	1	0	1	0	1
[15,]	-1	1	1	1	-1	-1	0	-1	0	1
[16,]	1	1	1	1	1	1	0	1	0	1
[17,]	2	0	0	0	0	2	0	2	0	0
[18,]	-2	0	0	0	0	-2	0	-2	0	0
[19,]	0	2	0	0	0	0	0	0	0	0

[20,]	0	-2	0	0	0	0	0	0	0	0
[21,]	0	0	2	0	0	0	0	0	0	2
[22,]	0	0	-2	0	0	0	-1	0	-1	-2
[23,]	0	0	0	2	0	0	1	0	1	0
[24,]	0	0	0	-2	0	0	1	0	1	0
[25,]	0	0	0	0	2	0	-1	0	-1	0
[26,]	0	0	0	0	-2	0	-1	0	-1	0
[27,]	0	0	0	0	0	0	1	0	1	0
[28,]	0	0	0	0	0	0	1	0	1	0
[29,]	0	0	0	0	0	0	0	0	0	0
[30,]	0	0	0	0	0	0	0	0	0	0

References

Allen, T. T., Yu, L., and Schmitz, J. (2003). An Experimental Design Criterion for Minimizing Meta-Model Prediction Errors Applied to Die Casting Process Design. *Journal of the Royal Statistical Society, Ser. C*, 52, 103–117.

Box, G. E. P. and Hunter, J. S. (1961). The 2^{k-p} Fractional Factorial Designs Part I. *Technometrics*, 3, 311–351.

Box, G. E. P. and Wilson, K. B. (1951). On the Experimental Attainment of Optimum Conditions. *Journal of the Royal Statistical Society,* Series B, 13, 1–45.

Bursztyn, D. and Steinberg, D. M. (2006). Comparisons of Designs for Computer Experiments. *Journal of Statistical Planning and Inference*, 136, 1103–1119.

Cao, Y., Smucker, B.J. and Robinson, T.J. (2017). A Hybrid Elitist Pareto-based Coordinate Exchange Algorithm for Constructing Multi-criteria Optimal Experimental Designs. *Stat Comput*. 27, 423–437.

Chang, S. (1997). An Algorithm to Generate Near *D*-optimal Designs for Multiple Response Surface Models. *IIE Transactions*, 29, 1073–1081.

Cooray-Wijesinha, M. and Khuri, A.I. (1987). The Sequential Generation of Multiresponse D-optimal Designs When the Variance–Covariance Matrix is Not Known. *Communications in Statistics—Simulation and Computation*, 16, 239–259.

Derringer, G. and Suich, R. (1980). Simultaneous Optimization of Several Response Variable. *Journal of Quality Technology*, 12, 214–219.

Draper, N. R. and Guttman, I. (1992). Treating Bias as Variance for Experimental Design Purposes. *Annals of the Institute of Statistical Mathematics*, 44, 659–671.

DuMouchel, W. and Jones, B. (1994). A Simple Bayesian Modification of D-Optimal Designs to Reduce Dependence on an Assumed Model. *Technometrics*, 36, 37–47.

Fedorov, V.V. (1972). Theory of Optimal Experiments. *New York: Academic Press*.

Gronwald, W., Hohm, T. and Hoffmann, D. (2008). Evolutionary Pareto- Optimization of Stably Folding Peptides. *BMC-Bioinformatics*, 9, 109.

Harrington, J. (1965). The Desirability Function. *Industrial Quality Control*, 21, 494–498.

Holland, J. H. (1975). Adaptation in Natural and Artificial Systems, Ann Arbor. *The University of Michigan Press*.

Kirkpatrick, S., Gelatt, C. and Vecchi, M. (1983). Optimization by Simulated Annealing. *Science*, 220, 671–680.

Jin, R., Chen, W. and Sudjianto, A. (2003). An Efficient Algorithm for Constructing Optimal Design of Computer Experiments. *Proceedings of DETC.03 ASME 2003 Design Engineering Technical Conferences and Computers and Information in Engineering Conference.*

Jones, B. and Nachtsheim, C. J. (2011). A Class of Three-Level Designs for Definitive Screening in the Presence of Second-Order Effects. *Journal of Quality Technology*, 43 (1), 1–15.

Lewis, R. M., Torczon, V., and Trosset, M.W. (2000). Direct Search Methods: Then and Now. *Journal of Computational and Applied Mathematics*, 124, 191–207.

Lu, L., Anderson-Cook, C.M. and Robinson, T.J. (2011). Optimization of Designed Experiments Based on Multiple Criteria Utilizing a Pareto Frontier. *Technometrics*. 53(4), 353–365.

Marget, W.M. and Morris, M.D. (2019). Central Composite Experimental Designs for Multiple Responses with Different Models. *Technometrics*, vol 61, issue 4.

Marler, R.T. and Arora, J.S. (2004). Survey of Multi-Objective Optimization Methods for Engineering. *Structural and Multidisciplinary Optimization*, 26, 369–395.

Myers, R. H., Montgomery, D. C. and Anderson-Cook, C. M. (2016). Response Surface Methodology: Process and Product Optimization Using Designed Experiments (4th ed.). *Wiley Series in Probability and Statistics*.

Nguyen, N. K. and Miller, A. J. (1992). A Review of Some Exchange Algorithms for Constructing Discrete D-optimal Designs. *Computational Statistics and Data Analysis*, 14, 489–498.

Park, Y.J. (2009). Multi-optimal designs for second-order response surface model. *Commun. Korean Statist. Soc*. 16 (1), 195–208.

Rady, E.A., Abd El-Monsef, M.M.E. and Seyam, M.M. (2009). Relationships among Several Optimality Criteria. *Interstat.statjournals.net*.

Sambo, F., Borrotti, M. and Mylona, K. (2014). A Coordinate-Exchange Two-phase Local Search Algorithm for the D- and I-optimal designs of split-plot experiments. *Comput. Statist. Data Anal*. 71, 1193–1207.

Trautmann, H. and Mehnen, J. (2009). Preference-Based Pareto Optimization in Certain and Noisy Environments. *Engineering Optimization*, 41, 23–38.