



I-optimal or *G*-optimal: Do we have to choose?

Stephen J. Walsh, Lu Lu & Christine M. Anderson-Cook

To cite this article: Stephen J. Walsh, Lu Lu & Christine M. Anderson-Cook (2023): *I*-optimal or *G*-optimal: Do we have to choose?, Quality Engineering, DOI: [10.1080/08982112.2023.2194963](https://doi.org/10.1080/08982112.2023.2194963)

To link to this article: <https://doi.org/10.1080/08982112.2023.2194963>



© 2023 The Author(s). Published with license by Taylor & Francis Group, LLC



[View supplementary material](#)



Published online: 13 Apr 2023.



[Submit your article to this journal](#)



Article views: 1510






[View related articles](#)



[View Crossmark data](#)

I-optimal or *G*-optimal: Do we have to choose?

Stephen J. Walsh^a , Lu Lu^b , and Christine M. Anderson-Cook^c 

^aDepartment of Mathematics and Statistics, Utah State University, Logan, UT, USA; ^bDepartment of Mathematics and Statistics, University of South Florida, Tampa, FL, USA; ^cLos Alamos, NM, USA

ABSTRACT

When optimizing an experimental design for good prediction performance based on an assumed second order response surface model, it is common to focus on a single optimality criterion, either *G*-optimality, for best worst-case prediction precision, or *I*-optimality, for best average prediction precision. In this article, we illustrate how using particle swarm optimization to construct a Pareto front of non-dominated designs that balance these two criteria yields some highly desirable results. In most scenarios, there are designs that simultaneously perform well for both criteria. Seeing alternative designs that vary how they balance the performance of *G*- and *I*-efficiency provides experimenters with choices that allow selection of a better match for their study objectives. We provide an extensive repository of Pareto fronts with designs for 17 common experimental scenarios for 2 (design size $N=6$ to 12), 3 ($N=10$ to 16) and 4 ($N=15, 17, 20$) experimental factors. These, when combined with a detailed strategy for how to efficiently analyze, assess, and select between alternatives, provide the reader with the tools to select the ideal design with a tailored balance between *G*- and *I*-optimality for their own experimental situations.

KEYWORDS



continuous designs; desirability function; exact designs; *G*-optimality; *I*-optimality; optimal experimental design; particle swarm optimization; prediction variance; Pareto front


1. Introduction

For response surface designs, experimenters frequently assume that a second-order model will adequately describe the underlying relationship between the response and experimental factors, and the choice of which designed experiment to run focuses on the ability to predict well throughout the input space. Assuming a particular model, the prediction variance (PV) – how much variability or uncertainty is associated with the prediction of the mean response value – can be calculated for the design at any point of prediction in the input space. Since the PV is a function of the natural variability of the process, which is typically unknown at the point of designing the experiment, we focus on the relative prediction variance (RPV), which removes this constant and allows easy comparisons between designs before data are collected. Historically, researchers have focused on either *I*-optimality (minimizing the average RPV across the input space) or *G*-optimality (minimizing the worst-case RPV) design criterion for this scenario. Both

criteria have some intuitive and practical appeal. The RPV throughout the input space represents a distribution of values, with statisticians and scientists often focusing on the center of distributions for a simple summary of their characteristics. Alternately, since one may not know *a priori* all locations where future predictions are needed, protecting against the worst possible RPV is appealing to bound the prediction uncertainty from the estimated model.

Algorithms, like the coordinate exchange (Meyer and Nachtsheim 1995), for generating optimal designs exist in most major statistical software packages. The *D*- and *A*-optimality criteria are commonly used when the focus is on model parameter estimation (such as in screening experiments), and *I*-optimality has relatively recently emerged as the dominant choice when prediction is the primary post-experiment analysis objective. The preference to use *I*-optimality over *G*-optimality for generating designs well suited for prediction appears to be influenced by two factors – (1) the tendency of statisticians to focus more frequently on

CONTACT Stephen J. Walsh  steve.walsh@usu.edu  Department of Mathematics and Statistics, Utah State University, Logan 84322, UT, USA.

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/08982112.2023.2194963>.

© 2023 The Author(s). Published with license by Taylor & Francis Group, LLC

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

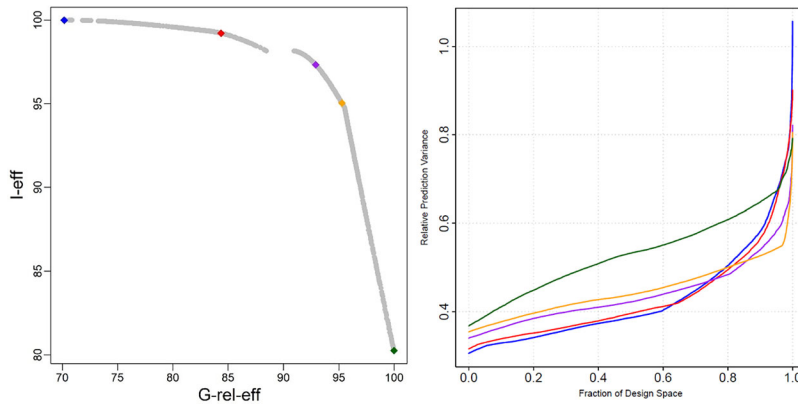


Figure 1. Summary of non-dominant designs for $K=2$, $N=9$ scenario. (a) the Pareto front with the I -optimal (blue) and G -optimal (green) designs. (b) the FDS plot of the 5 highlighted designs marked in (a).

the central characteristic of distributions, here the RPV throughout the input space, and (2) the computational ease of generating I -optimal designs compared with generating G -optimal designs (Rodriguez et al. 2010; Walsh and Borkowski, 2022). But should experimenters be automatically guided to I -optimality when they want to focus on the quality of prediction of a design? Are there benefits to seeing and comparing the I - and G -optimal designs to understand how well each does relative to the other criterion? And, perhaps most importantly, are there designs that might be more appealing to practitioners than either the I -optimal or G -optimal designs? For example, there may exist designs that retain high performance on both criteria.

After exploring a suite of rational alternatives that balance the two RPV-based optimality criteria, we believe that for many experimental scenarios, there are superior choices available other than the strictly optimal design for either G - or I -criterion. Focusing on generating a good design under only one of the criteria potentially ignores the other completely. In many cases, a small reduction in the efficiency of one criterion can lead to disproportionate improvement in the other. The efficiency of a design based on a criterion (formally defined in Section 2.1) ranges from 0 to 100% and describes how well the design does relative to the best possible design for that criterion.

Consider the design scenario with 2 continuous input factors and an experiment with 9 experimental runs in a rectangular study region. The algorithm for finding the I -optimal design does not consider G -efficiency at all. Likewise, the G -optimal design was constructed with focus only on that criterion in isolation. Figure 1(a) shows a scatterplot of the I - and G -efficiencies of these designs plus the suite of rational alternatives found on the Pareto front, that is, the set of non-dominated designs based on the two criteria. From this plot, we see that the I -optimal design (blue dot) that is by definition

100% I -efficient is only 70.2% as G -efficient as the G -optimal design. This means that the worst-case RPV over the input space is $(1/0.702 - 1) = 0.425$ or 42.5% larger than the worst case from the G -optimal design. Similarly, the G -optimal design (green dot with 100% G -efficiency) is only 80.2% as I -efficient as the best available. This means that the average RPV for this design is $(1/0.802 - 1) = 0.247$ or 24.7% larger than the best possible average. Thus, focusing exclusively on one criterion has led to mediocre performance on the other. However, as we advocate in this article, if both criteria are considered during the design construction phase, then we can construct some attractive designs that have the following efficiencies:

$$(I_{\text{eff}}, G_{\text{rel-eff}}) = \begin{array}{l} (99.2\%, 84.4\%) \text{ (red dot)} \\ (97.3\%, 92.9\%) \text{ (purple)} \\ (95.0\%, 95.3\%) \text{ (orange)}. \end{array}$$

If we consider the first of these three options, this means that there is a design that has an average RPV that is only $(1/0.992 - 1) = 0.008$ or 0.8% larger than the I -optimal design and has a maximum RPV that is $(1/0.844 - 1) = 0.185$ or 18.5% larger than the G -optimal design. Similarly, for the third option, the design has both the average and maximum RPV that is only $\sim 5\%$ larger than the best possible for each criterion.

Figure 1b shows the fraction of design space (FDS) plots (Zahran, Anderson-Cook, and Myers 2003) for these 5 designs. FDS plots show a summary of the entire range of RPV values for a given design in the input space defined by the range of the experiment settings. The vertical-axis shows the RPV values, while the horizontal-axis shows the proportion of the input space at or below a particular RPV value for that design. For example, for the I -optimal design (blue), the point $(0.6, 0.4)$ means that 60% of the inputs space has RPV at or below 0.4. The ideal design has an FDS curve that is as low and flat as possible, as this

indicates good precision in prediction throughout the input space. For *I*-optimality, the focus is on having the average value (generally found near the center of the horizontal-axis range) of the line as small possible. Note how the blue curve has the lowest values for most of the horizontal-axis range, but at the expense of a very large maximum value. *G*-optimality seeks to make the right endpoint of the curve (the maximum RPV) as small as possible. The associated green curve has the lowest maximum, but at the expense of considerably larger values across most of the horizontal-axis range. The red, purple, and orange curves from the Pareto front highlight the potential of more balanced solutions. They are each able to lower the RPV values throughout the horizontal-axis range, relative to the green *G*-optimal curve, while simultaneously lowering the maximum (worst-case) RPV value relative to the blue *I*-optimal curve. For many experimenters, the middle 3 curves highlight solutions that may represent better alternatives to either of the optimal designs. The key is to be able to construct these designs through an effective optimization algorithm, and then be able to examine the alternatives to identify the one that is the best match to experimental goals.

In this article, we explore 17 design scenarios for $K=2, 3$, and 4-factor designed experiments with the goal of good prediction precision for a standard second-order response surface model. We present the two optimal designs based on either criterion alone, and then provide a rich set of alternative designs on the Pareto front (PF) based on the two criteria that allow the practitioner to choose the right balance for their experiment. For some experiments, it is critical to protect against poor worst-case prediction. In other cases, it may be more appealing to choose a design that performs well for the average RPV. In most cases, designs exist that perform extremely well, but not strictly best, for both criteria simultaneously. As a key contribution of this article, we provide a repository of designs which characterize the PFs for the 17 design scenarios considered in the [Supplementary Materials](#) (SM). Part 1 of the SM describes the spreadsheet containing all of the designs for each of the PFs. R-code has been included that allows for easy extraction of any selected design. With this repository, the reader is equipped with the tools needed to select the best design for their experimental situation that finds the appropriate balance between the worst-case and average RPV throughout the input space.

Regardless of which design is finally selected, we argue that knowledge is power: understanding the

available alternatives enables clearer comparisons and ultimately better decision-making when selecting the design to implement in practice. In addition, the trade-offs between the *G*- and *I*-criteria differ substantially depending on the number of input factors and the size of the design sought. Seeing what is possible for a particular experimental scenario provides valuable information, and we propose a structured process for carefully considering different choices and comparing their performance.

The remainder of the article is organized as follows: [Sections 2.1–2.3](#), respectively, provide background on the RPV-based optimality criteria, Pareto fronts (PF) and the particle swarm optimization (PSO) algorithm used to identify designs worthy of further consideration for each experimental scenario. [Section 3](#) illustrates the complete method for comparing *I*- and *G*-optimal designs to rational alternatives that balance the two objectives and a process for selecting the right design for each experimenter's preferred compromise between objectives. [Section 4](#) shows the suites of optimal designs and Pareto fronts for several different sized designs with 2 continuous input factors. [Section 5](#) summarizes results for designs with 3 and 4 continuous input factors. Finally, [Section 6](#) contains conclusions. The SM include (1) a 5-part document with additional information, (2) the repository containing 17 .csv files labeled "ParetoFront_K#_N*.csv", where "#" indicates the number of continuous factors, and "*" indicates the design size, and (3) three files with R code to easily extract designs from the repository and implement the design selection methods described in [Section 3](#).

2. Background

This section provides background on the components needed to implement the methods described in the remainder of the article: the RPV-based optimality criteria, Pareto fronts (PF) and the particle swarm optimization (PSO) algorithm used to identify worthy designs.

2.1. Prediction variance based optimality

We now introduce some of the background of the design metrics considered in this article as well as a historical perspective on the optimality criteria for prediction.

Second-order models are a common choice for locally approximating a wide variety of smooth and continuous underlying relationships between a set of K experimental factors and a response. They provide a robust approximation since they can be thought of as a parameterized version of the second-order Taylor

series expansion (Myers, Montgomery, and Anderson-Cook 2016). The assumed model form is

$$y = \beta_0 + \sum_{i=1}^K \beta_i x_i + \sum_{i=1}^{K-1} \sum_{j=i+1}^K \beta_{ij} x_i x_j + \sum_{i=1}^K \beta_{ii} x_i^2 + \varepsilon$$

and includes an intercept, K main effects, $K(K-1)/2$ two-way interactions and K quadratic terms. The error term, ε , is assumed to be independently and identically distributed as $\mathcal{N}(0, \sigma^2)$. We consider the scenario for an experiment with N design points and K experimental factors. For the results summarized in this article, we assume all design factors are scaled to range $[-1, 1]$ and so the design space is the $\chi = [-1, 1]^K$ hypercube, where all settings of each input in that range are assumed possible. In the conclusions we discuss how this assumption may be relaxed to include categorical design factors. Hence, we are considering the class of continuous designs. A design point (i.e., an experimental run) is an $\mathbf{x}' : 1 \times K$ row-vector of the form, $\mathbf{x}' = (x_1 \ x_2 \ \dots \ x_K)$ and specifies the values of all inputs to be implemented for that experimental run. We use $\mathbf{X} = [\mathbf{x}_1 \ \mathbf{x}_2 \ \dots \ \mathbf{x}_N]'$ to denote the $N \times K$ design matrix with the rows representing the N design points.

For estimation of the model and for prediction of new observations in the input space, let \mathbf{F} be the $N \times p$ model matrix with rows given by the expansion vector

$$\mathbf{f}'(\mathbf{x}'_i) = (1 \ x_{i1} \ \dots \ x_{iK} \ x_{i1}x_{i2} \ \dots \ x_{i(K-1)}x_{iK} \ x_{i1}^2 \ \dots \ x_{iK}^2).$$

The model can be written in matrix-vector form as $\mathbf{y} = \mathbf{F}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. The ordinary least squares estimator of $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}} = (\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\mathbf{y}$, which has variance $\text{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{F}'\mathbf{F})^{-1}$, where σ^2 is the variance of $\boldsymbol{\varepsilon}$. The total model information matrix for $\boldsymbol{\beta}$, specifically $\mathbf{M}(\mathbf{X}) = \mathbf{F}'\mathbf{F}$, plays an important role in optimal design of experiments as all commonly used optimal design objective functions are functions of this matrix.

To implement the construction of an ideal experimental design in practice, an optimality criterion is selected and used to define which candidate designs $\mathbf{X} \subset \chi^N$ are highly desirable. An optimization algorithm is required to search χ^N to find the “best,” or optimal, design. Thus, an exact optimal design problem (Goos and Jones 2011) is defined by the three components:

- (1) The number of design points N affordable in the experiment.
- (2) The structure of the model one wishes to fit (here, the second-order model).

- (3) A criterion which defines an optimal design. This is a function of $\mathbf{M}(\mathbf{X})$.

In this article we consider the G - and I - criteria described below. Both criteria are functions of the variance of a mean prediction, $\hat{y}(\mathbf{x}')$, at an arbitrary location $\mathbf{x}' \in \chi$, which is defined as

$$\text{Var}(\hat{y}(\mathbf{x}')) = \sigma^2 \mathbf{f}'(\mathbf{x}')(\mathbf{F}'\mathbf{F})^{-1} \mathbf{f}(\mathbf{x}').$$

Since the natural variability of the experimental process, σ^2 , is generally unknown at the time of planning the experiment, we consider the RPV which removes the scale parameter σ^2 by multiplying by the factor $1/\sigma^2$. The RPV (of a candidate design \mathbf{X}) at any location $\mathbf{x}' \in \chi$, is calculated as

$$\text{RPV}(\mathbf{x}') = \mathbf{f}'(\mathbf{x}')\mathbf{M}(\mathbf{X})^{-1}\mathbf{f}(\mathbf{x}') = \mathbf{f}'(\mathbf{x}')(\mathbf{F}'\mathbf{F})^{-1}\mathbf{f}(\mathbf{x}')$$

and allows for easy scale-free comparisons between designs of the same or different sizes (Jensen 2018).

The G -score of an arbitrary design \mathbf{X} measures the maximum RPV compared to the theoretical optimum over all points of prediction in $\mathbf{x}' \in \chi$. There is a known theoretical lower bound for the RPV maximum of $p/N = \left(1 + 2K + \frac{K(K-1)}{2}\right)/N$, where p is the number of parameters in the second order model (Myers, Montgomery, and Anderson-Cook 2016). Hence, the G -score (often called G -efficiency) is defined as $G(\mathbf{X}) = 100 \frac{p/N}{\max_{\mathbf{x}' \in \chi} \text{RPV}(\mathbf{x}')}$, with larger values indicating smaller maximum RPV, i.e., more precise worst-case prediction. The G -optimal design, \mathbf{X}^* , is the design which maximizes, over all designs $\mathbf{X} \subset \chi^N$, the G -score.

Evaluating the G -score for candidate designs remains a computational challenge because the score itself is defined as an optimization throughout the design space,

$$G(\mathbf{X}) = \frac{p}{N \max_{\mathbf{x}' \in \chi} \text{RPV}(\mathbf{x}')}.$$

This calculation has proved troublesome for global optimal design searches, see Walsh and Borkowski (2022) for some discussion on this topic. Recognizing a symmetry in the RPV surface for second order models, several authors have found an approximation to this calculation to be adequate and give small error when calculating the G -criterion for a candidate design (see, e.g., Borkowski 2003; Hernandez and Nachtsheim 2018; Walsh 2021; Walsh and Borkowski 2022). These authors employ a 5^K grid over χ with the grid defined as $G_\chi = \{-1, -0.5, 0, 0.5, 1\}^K$ to

calculate RPV with the maximum value of RPV over these points taken as an approximation to the G -score.

The I -criterion is sometimes referred to as an integrated variance criterion. An I -optimal design minimizes the average RPV over the design space, χ . Hence, we can think of the I -optimal design, \mathbf{X}° , as that which minimizes the average relative prediction variance $ARPV(\mathbf{X}) = \text{avg}(RPV(\mathbf{x}'))$, over all designs $\mathbf{X} \subset \chi^N$. In this article, for a candidate design \mathbf{X} , we compute an I -score as the inverse of the I -criterion (so larger I -score means better design)

$$I(\mathbf{X}) = \frac{1}{ARPV(\mathbf{X})} = \frac{V}{\int \mathbf{f}'(\mathbf{x}')(\mathbf{F}'\mathbf{F})^{-1}\mathbf{f}(\mathbf{x}')\mathbf{d}\mathbf{x}'}, \text{ or } I(\mathbf{X}) = \frac{V}{\text{tr}\{(\mathbf{F}'\mathbf{F})^{-1}\mathbf{W}\}},$$

where $\text{tr}\{ \cdot \}$ is the trace operator, $\mathbf{W} = \int \mathbf{f}(\mathbf{x}')\mathbf{f}'(\mathbf{x}')\mathbf{d}\mathbf{x}'$ is referred to as the region-moments matrix, and V represents the volume of the design space. A larger I -score indicates more precise prediction averaged cross the design space. Provided the design space is a simple geometry, such as the K -dimensional hypercube (as in the problems addressed in this article), the multi-dimensional integral has closed-form and can be computed exactly (see Myers, Montgomery, and Anderson-Cook 2016).

In this article, we consider designs that are not strictly the G - or I -optimal designs, but rather seek good performance on both criteria. Hence, we define the relative efficiency for a given design, \mathbf{X} , as

$$G_{rel-eff} = 100 \frac{G(\mathbf{X})}{G(\mathbf{X}^*)} \text{ and } I_{eff} = 100 \frac{I(\mathbf{X})}{I(\mathbf{X}^\circ)},$$

where the relative efficiencies lie in $[0,100]\%$ and the optimal design is in the denominator for each equation as we seek to maximize both the G - and I -scores.

Historically, both G - and I -optimality in the context of exact designs, were the subject of much academic research, with I -optimality appearing to have gained more widespread application among practitioners. Borkowski (2003) applied a genetic algorithm to generate both G - and I -optimal designs for the second-order model and twenty-one exact small response surface design scenarios covering $K=1, 2, 3$ design factors and experiment sizes $N=3, \dots, 9$, $N=6, \dots, 12$, and $N=10, \dots, 16$, respectively. Rodriguez et al. (2010) developed an augmented coordinate exchange algorithm to reproduce the designs generated by Borkowski (2003) and extended the study to $K=4, 5$ factors and design sizes $N=15, 20, 24$ and $N=21, 23, 30$, respectively.

Rodriguez et al. (2010) focused specifically on generating G -optimal designs and comparing the proposed G -optimal design's prediction variance properties to those of corresponding I -optimal designs. They noted that, while G -optimal designs have lower maximum RPV, the I -optimal designs had lower RPV over a larger portion (say $> 80\%$) of the design space. This observation, in conjunction with the fact that it is considerably more difficult to generate G -optimal designs than I -optimal designs, led the authors to recommend that, from a practitioner's perspective, the added expense and difficulty of generating G -optimal designs may not be worth the effort nor risk (of not finding the optimal design). Hernandez and Nachtsheim (2018) published a new augmented coordinate exchange algorithm based on continuous I -optimal preference designs as a starting point, to generate highly G -optimal designs in a reasonably short computing time. They covered the Borkowski (2003) scenarios and proposed G -optimal designs for scenarios $K=4, N=17$, and $K=5, N=23$. Most recently Walsh and Borkowski (2022) provide a detailed discussion on the history and difficulty of generating G -optimal designs and provide the current best-known G -optimal designs for the design scenarios discussed in this article generated efficiently via an adaptation of particle swarm optimization.

2.2. Pareto fronts and multi-objective optimal design

Next, we introduce some background for Pareto fronts, which are used extensively in this article to find rational solutions that strike different balances between the G - and I -criteria. The Pareto front (PF) has been broadly used for multi-objective optimization (Chen and Zhou 2022; Lu, Li, and Anderson-Cook 2016; Lu, Anderson-Cook, and Zhang 2021; Fox et al. 2019; Trautmann and Mehnen 2009; Gronwald, Hohm, and Hoffmann 2008) and has recently been adapted for design optimization (Lu, Anderson-Cook, and Robinson 2011; Lu and Anderson-Cook 2014, 2021). The method finds the set of non-dominated solutions based on multiple criteria, and hence extracts a set of objectively superior solutions before considering the subjective user priorities. Consider our problem of maximizing G - and I -efficiencies. A design is said to Pareto dominate another if G - and I -efficiencies of the first design are as large as those of the latter design and at least one of the G - or I -efficiency of the first design is strictly larger than the criterion value of the latter design. Hence, the PF contains the designs that are not outperformed by any

other design based on all the criteria (here two) under consideration.

To find the PF from a collection of M designs with calculated G - and I -efficiencies, the following steps can be used to identify the PF:

1. Add the first design from the collection into the current PF.
2. Update the current PF with the i design in the collection via steps 2a-c and repeat this for $i = 2, \dots, M$.
 - 2a. If the i th design is Pareto dominated by at least one design on the current PF, discard the i th design without changing the current PF;
 - 2b. If the i th design Pareto dominates at least one design on the current PF, remove the dominated design(s) and add the i th design to the current PF;
 - 2c. If the i th design does not Pareto dominate any design and is not dominated by any design on the current PF, then add the i th design to the current PF without removing any other designs.

Note the above steps can be used not only to find a PF from a finite set of designs, but also to be built into a search algorithm for continuing to populate the PF.

The goal of the proposed methodology is to provide a suite of multiple designs, which are not strictly optimal for either prediction variance criteria, but rather that simultaneously achieve good performance for both. The degree to which the chosen design achieves near-ideal performance involves considering the trade-offs between the available alternatives and matching the experimental priorities to what designs are available. The relative emphasis that is placed on performance of G - and I -optimality is a subjective choice that the experimenter can make after examining the set of non-dominated PF solutions.

To find suitable designs to compare and consider, a PF is constructed to find all the non-dominated designs for the G - and I -optimality criteria. This builds on the algorithms developed for efficiently creating a suite of designs that trade-off their level of performance on the set of user-specified criteria. In design optimization, point exchange and coordinate exchange algorithms have been broadly used with pre-defined candidate sets for finding exact optimal designs on the PF (Lu, Anderson-Cook, and Robinson 2011; Lu and Anderson-Cook 2013; Cao, Smucker, and Robinson 2017). In addition, multi-objective

evolutionary algorithms have been popular for finding the PF solutions for broad applications (Deb et al. 2002; Deb and Jain 2014; Lu, Chapman, and Anderson-Cook 2013; Chapman, Lu, and Anderson-Cook 2018). Recently, particle swarm optimization (PSO), has gained popularity in design optimization. See Chen, Chen, and Wong (2022) and Walsh and Borkowski (2022) for a comprehensive review. In this article, we explore the use of PSO for populating the PF of non-dominated designs based on G - and I -optimality criteria.

Note that the identification of the PF is solely based on the criteria values of individual solutions and does not depend on any subjective choices. Hence, it has been broadly accepted that identifying the PF offers a good first step for providing an objectively best set of alternatives from which the experimenter can choose based on other subjective factors. Using exploratory and graphical methods, described in Section 3, the design that best achieves the desired balance between good G - and I - performance to match the experimental objectives can be identified.

2.3. Particle swarm optimization algorithm

To construct the PF based on the two prediction variance criteria, we use an adaptation of particle swarm optimization (PSO) algorithm as discussed in Walsh and Borkowski (2022). PSO is a fast and effective meta-heuristic optimization algorithm well-suited for generating highly optimal exact designs. Several particles or agents comprise the swarm that explores the solution space looking for the optimum. Each particle tracks its personal best solution, while the overall algorithm tracks the global best. Evolution of the swarm considers both the personal and global or groups best and balances these pieces of information to control how particles step through the search space in search of the optimal solution.

PSO is an attractive optimizer for the optimal design problem because 1) it makes little to no assumptions regarding the structure of the objective function, 2) it is non-greedy, allowing all elements of the design to change during the search and 3) it is robust to entrapment to local optima. The combination of these characteristics leads to a high probability of finding the global optimum in a single run relative to local optimizers like the coordinate or point exchange algorithms. Optimal design problems are well-known to have high-dimensional multimodal objectives (Lin et al. 2015) and thereby are a good match to the PSO algorithm.

PSO is a relatively new tool for solving the optimal design problem. Chen et al. (2014) were the first to use PSO for the construction of optimal designs and generated Latin-hypercube designs. Chen et al. (2014) and Mak and Joseph (2018) applied PSO to generate space-filling designs. PSO for generating optimal designs for non-linear models were proposed in Lukemire, Mandal, and Wong (2019), Chen et al. (2015), and Chen et al. (2019). PSO-generation of Bayesian continuous designs is discussed in Shi, Zhang, and Wong (2019) and a PSO algorithm for constructing continuous optimal designs for mixture experiments is presented in Wong et al. (2015). The literature is relatively lacking for applications and adaptations of PSO to generate exact optimal designs, but a comprehensive review and development is provided by Walsh (2021).

For the full details of the implementation of PSO applied to the searches in this article, see Walsh and Borkowski (2022). Here we provide a brief overview of the key parts of the search algorithm. The goal of the search is to optimize a scalar criterion. PSO proceeds by randomly generating a number, S , candidate designs which form a swarm, and then employs the following update equations which cause the “design swarm” to step through the space of candidate matrices in search of the optimal design:

$$\begin{aligned} \mathbf{V}_i(t+1) &= \omega \mathbf{V}_i(t) \text{ (inertia)} \\ &+ c_1 U_K(0, 1) \odot (\mathbf{P}_{\text{best}, i} - \mathbf{X}_i(t)) \text{ (cognitive)} \\ &+ c_2 U_K(0, 1) \odot (\mathbf{G}_{\text{best}} - \mathbf{X}_i(t)) \text{ (social)} \\ \mathbf{X}_i(t+1) &= \mathbf{X}_i(t) + \mathbf{V}_i(t+1) \text{ (design moves)} \end{aligned}$$

where $\mathbf{X}_i(t)$ represents candidate design i at time t , $\mathbf{V}_i(t)$ is the corresponding velocity which governs both step-size and direction, $\mathbf{P}_{\text{best}, i}$ represents design's personal best location found during the search, \mathbf{G}_{best} represents the groups best location found (this is the solution of the optimization search at the end of the algorithm run), \odot denotes Hadamard product, $U_K(0, 1)$ represents a K -dimensional uniform random vector with values drawn between $(0, 1)$, and ω , c_1 , and c_2 are the PSO velocity update scaling weights. The optimal values of these weights for searches in standard Euclidean spaces have been shown to be $\omega = \frac{1}{\log 2}$, and $c_1 = c_2 = \frac{1}{2} + \log 2$ (Clerc and Kennedy 2002; Clerc 2012) and so we adopt these values in this article.

The quality of each design as measured by the chosen criterion of interest is defined as its desirability (a weighted combination of G - and I -criteria) denoted by $\text{Des}(\mathbf{X}_i(t))$, which without loss of generality we

seek to maximize. As the S initially randomly selected designs in the swarm search the space of candidate designs for the global optimizer, the personal best and group's best designs found are updated via the simple logic:

$$\begin{aligned} &\text{if } \text{Des}(\mathbf{X}_i(t+1)) > \text{Des}(\mathbf{P}_{\text{best}, i}) \\ &\quad \mathbf{P}_{\text{best}, i} \leftarrow \mathbf{X}_i(t+1) \\ &\text{if } \text{Des}(\mathbf{P}_{\text{best}, i}) > \text{Des}(\mathbf{G}_{\text{best}}) \\ &\quad \mathbf{G}_{\text{best}} \leftarrow \mathbf{P}_{\text{best}, i} \\ &\text{end if} \\ &\text{end if.} \end{aligned}$$

While most applications of PSO for generating optimal designs have utilized the so-called global communication topology (see, e.g., Chen et al. 2014; Wong et al. 2015), Walsh and Borkowski (2022) recently demonstrated that the local communication topology yields significant increases to the probability that a single run of PSO would find the globally optimal design. The algorithm can be flexibly applied for any scalar criterion to be optimized. For the $K=2, 3$ and 4 scenarios explored, the standard PSO algorithm was initially applied for each of the G - and I -criteria separately. This provided calibration for the range of results possible, and defined clear ending points of the PF. To populate the remainder of the PF, adaptations to the algorithm to combine the criteria were needed.

To generate candidate G/I Pareto front optimal designs, we define how the PSO algorithm can be adapted to simultaneously handle the two criteria (G and I) for this scenario. We use desirability functions (Derringer and Suich 1980) to combine them into a single metric to guide the direction of our searches for all different emphases of the criteria. In particular, the additive desirability function provides a way to direct the optimization objective which is defined as

$$\text{Des}(\mathbf{X}|w) = wG_s(\mathbf{X}) + (1-w)I_s(\mathbf{X})$$

where G_s and I_s are scaled/calibrated values of the G and I -criteria following the approach of Lu, Anderson-Cook, and Robinson (2011). Designs defined by

$$\mathbf{X}^* = \text{argmax Des}(\mathbf{X}|w)$$

are promising designs and have strong potential to be Pareto-optimal designs given a selected w . To explore all regions of the Pareto front, we consider a set of weights, w in a regular sequence of 50 values in $(0.01, 0.03, \dots, 0.99)$ and apply PSO multiple times per w to generate Pareto optimal designs for locations on the PF close to each desirability weight, w . The final Pareto-optimal designs provided in the repository are

on the PF of the combined designs over the set of weights explored.

This approach provides an effective way to explore and populate and approximate the entire range of the PF. It remains an open research question about how to further improve the efficiency of PF generation and approximation. We are currently exploring alternate strategies to speed up the search and provide even more complete PFs as part of continuing research. The set-ups of the optimization run campaigns and PSO parameters used to generate the databases from which we extracted G/I -PFs for each $\{N, K\}$ design scenario are provided in Part 2 of the SM.

3. Example $K = 2, N = 9$: Design selection from a Rich Pareto front

The overall process of selecting the ideal design for a particular scenario involves several stages:

1. Identify the criteria on which to base the selection.
2. Construct a Pareto front of non-dominated solutions based on the chosen criteria.
3. Select the best design from among the PF choices that best matches the particular priorities and goals of the experiment.

In this article, the first two stages have been provided for the reader when the emphasis is on good prediction variance performance. The G - and I -optimality criteria were used to construct the PFs using PSO with choices for each of the 17 design scenarios described. This section focuses on how to effectively and efficiently complete the third stage where the experimenter chooses the ideal design from a rich PF for their requirements. R-code is included in the SM that allows straightforward implementation of Step 3 described below.

Returning to the example in Figure 1, we see that for the 2-factor case with 9 experimental runs, there are designs that balance G - and I -optimality criteria and offer good performance for both. So, how would an experimenter proceed from having a rich Pareto front that visually offers multiple alternatives that look promising, to making a final choice of the one design to be implemented and run?

The PF offers an objective (not subjective) set of desirable solutions. However, a process for moving from many solutions to selecting a single final design from the PF of the non-dominated solutions requires an intentional plan with supporting tools. A rich set of numerical and graphical summaries (Lu, Anderson-

Cook, and Robinson 2011; Lu, Chapman, and Anderson-Cook 2013; Lu, Chapman, and Anderson-Cook 2017; Lu and Anderson-Cook 2012) have been proposed for further examining PF solutions and their tradeoffs to facilitate informed solution selection. As the richness of the PF solutions increases, interpreting results and making further decisions becomes more challenging even with the aid of these graphical and analytical tools. This is particularly true when we consider design selection over a continuous input space. The identified PF in Figure 1a is so densely populated that it is close to continuous. To make solution selection more manageable, we suggest a process to select a final design from a rich PF, which includes 1) using an efficient thinning approach to select a subset of representative solutions from the dense PF and 2) choosing a design from the reduced set with numerical and graphical summaries to best match experimental goals.

We use the $K = 2$ and $N = 9$ case from Figure 1a to illustrate this structured design selection process based on the dense PF in the provided repository. The Pareto front contains 2508 designs and hence looks almost continuous for most portions. There are a few small gaps on the PF due to discontinuities in the desirability function. The shape of the PF is convex up toward the Utopia Point (UP) which is the theoretical optimum with the best values of both criteria and is generally unobtainable as a real solution when considering both optimality criteria. The shape of the PF indicates there is a less severe tradeoff between the two criteria and one is more likely to be able to find solutions that offer acceptable performance on both criteria. On the other hand, if the shape of the PF is concave away from the UP, then it suggests more severe tradeoffs and an improvement on one criterion is likely to require more sacrifice on the performance of another criterion. A spreadsheet consisting of all the 2508 designs on the PF with their G - and I -efficiencies for the $K = 2$ and $N = 9$ case is in the SM repository.

3.1. Thinning a dense Pareto front

The proposed thinning process was adapted from ε -dominance proposed by Laumanns et al. (2002) to improve convergence and diversity of PF solutions for multi-objective evolutionary algorithms. The concept of ε -dominance was proposed as a relaxed Pareto dominance relationship to provide a smaller approximation of the full Pareto set (Hernandez-Diaz et al. 2007; Reuter 1990). There have been two kinds of ε -dominance commonly discussed: additive ε -dominance based on absolute

deviation and multiplicative ε -dominance based on relative deviation. We adapt the additive ε -dominance concept. Without loss of generality, we illustrate the concept with a maximization problem with K objectives. A solution x is said to ε -dominate another solution y for some $\varepsilon > 0$ and denoted as $x \succ_{\varepsilon} y$, if and only if $x_j + \varepsilon \geq y_j$ for $j = 1, \dots, K$. Hence, some non-dominated Pareto solutions can become ε -dominated under the relaxed dominance relationship, which allows for a thinner PF to be identified.

The method divides the criterion space into a set of hypercubes of length ε (for 2 criteria, this is a square). Then all the PF solutions located in the same hypercube are ε -dominated to each other. In this case, we can select a smaller set of solutions by taking only one solution from each hypercube of ε -dominated solutions while solutions from different hypercubes are non- ε -dominated. Then, a smaller set of non-dominated hypercubes consisting of PF solutions are selected to ensure the reduced solutions are more spread out across the criteria values. The two single criterion optimal designs are included in the thinned PF to ensure a complete range of the PF values. Below we summarize the key steps of the thinning process for a rich PF based on a chosen ε value.

Algorithm 1. Thin a rich PF based on ε -dominance

Step 1: Scale each criterion of PF solutions to $[0,1]$, where 1 corresponds to the best value and 0 to the worst value of all points on the PF;

Step 2: For a selected $\varepsilon > 0$ value, divide the PF criterion space into hypercubes of length ε ;

Step 3: Select the set of non-dominated hypercubes that contain PF solutions (many hypercubes will not contain any solutions from the PF);

Step 4: Select the solution from each hypercube that minimizes the Euclidean distance to the hypercube Utopia point (the theoretical best value within each hypercube, denoted by $(r \cdot \varepsilon, c \cdot \varepsilon)$ where r and c represent the row and column numbers of the hypercube from the origin);

Step 5: Add the single criterion optimal solutions to the thinned PF, if not already included in the earlier step.

Figure 2a shows an illustrative example of the thinned PF based on a chosen ε value for a hypothetical PF for a maximization problem with two criteria. The solid circles are the solutions included in the thinned PF while the open circles are not selected. The open circles have an adjacent square either above

or to the right of them containing a PF solution, and hence were not included in the thinned PF. The red dots present the single objective optimal solutions that are included in the thinned PF to ensure full coverage of the criteria values across the entire original PF. Considering that the number of reduced solutions depends on the choice of ε and the shape and richness of the original PF, we recommend exploring different ε values to select a desired density of the thinned PF for subsequent solution selection.

Figure 2b shows the thinned PF (shown as the red dots) for the full PF (shown in gray) from Figure 1a for the $K=2$ and $N=9$ case. The thinned PF contains 20 solutions and was obtained by using algorithm 1 with $\varepsilon = 0.02$. Recall that the original PF contains 2508 solutions, which is impractical to consider for individual examination. By using the proposed thinning approach, the rich set of PF solutions is effectively reduced to a small number of manageable solutions that evenly spread over the original PF and offers good coverage of the range of possible criterion values for each of G - and I -efficiency. Note a larger choice of ε would result in a thinner PF, while a smaller value of ε produces a denser PF.

3.2. Design selection from a reduced set

After we have thinned the PF to a smaller manageable set of solutions, the next step is to carefully examine the individual solutions to understand their performance and tradeoffs to match the solutions to user priorities. There are different methods for selecting the final solutions. In this subsection, we describe a few common options in multi-criterion optimization.

The simplest is the threshold approach which does not require the PF to be thinned before selecting a best design. In this case, we first identify a primary criterion, considered most important among the considered objectives. We determine a poorest acceptable value for this primary criterion. For example, with I_{eff} as the primary criterion and a lower bound for performance defined as 95%, the method selects the best G -efficiency design subject to the constraint on the acceptable I -efficiency. Hence the design with $(I_{\text{eff}}, G_{\text{rel-eff}}) = (95.0\%, 95.3\%)$ (shown as the orange point in Figure 3a) would be selected based on the threshold approach.

Another commonly used approach, which also does not require the preliminary thinning step, is the Utopia point approach (Marler and Arora 2004). The idea is to select a best solution based on minimizing its distance to the Utopia point, the theoretical

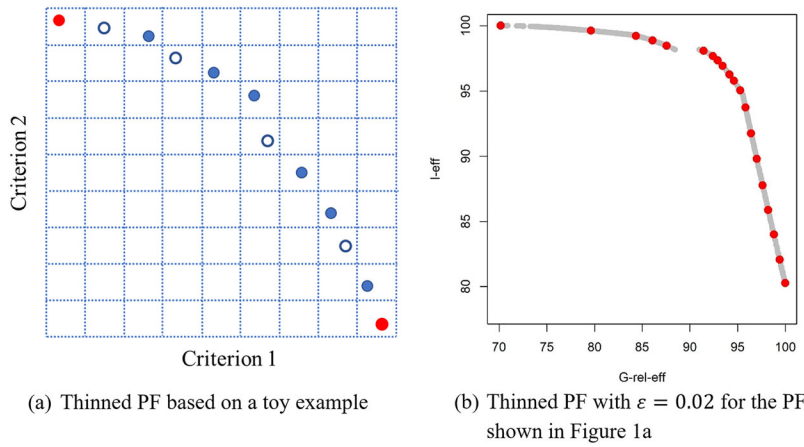


Figure 2. Illustration of thinned PF for (a) a toy example to demonstrate the concept and (b) the PF of 2508 solutions from Figure 1a for the case of $K = 2$ and $N = 9$.

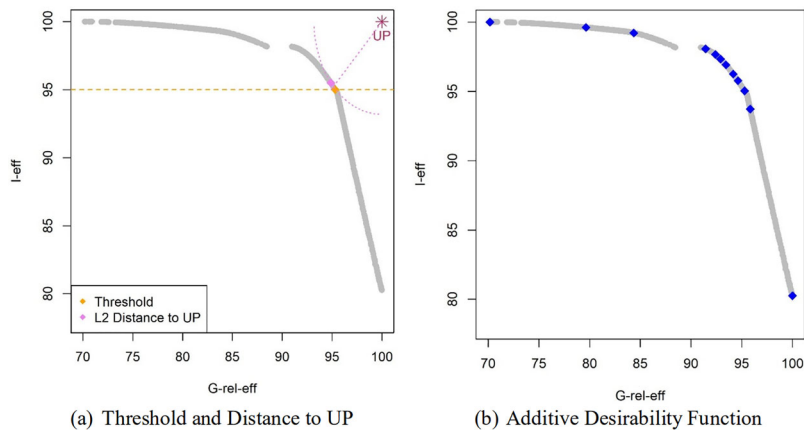


Figure 3. (a) Selected designs based on the threshold approach (orange point) and minimizing the Euclidean distance to the Utopia point (UP, violet point); (b) Selected designs based on the additive desirability function over 1000 evenly spread weight combinations overall the entire range of possible values.

solution (here, design) that simultaneously achieves the best values for all criteria. Different distance measures, such as L_p -norm for $p \geq 1$, could be used. One of the common choices is the Euclidean distance, i.e., the L_2 -norm. The violet point in Figure 3a with the criteria values at $(I_{\text{eff}}, G_{\text{rel-eff}}) = (95.5\%, 94.9\%)$ is the selected solution based on minimizing the Euclidean distance to the Utopia point (on the scaled $[0,1]$ ranges). Note the Utopia point approach implicitly assumes equal emphasis of both criteria under a chosen distance measure and hence may not offer as much control about the location of the point on the PF or the best match of desired user priority.

The third and more sophisticated approach that allows greater user control is to allow weights to represent different user priorities and their impact to be explicitly evaluated and considered for selecting the final solution. One way is to find optimal solutions for a fine grid of all possible weights based on a chosen desirability function (DF) (Derringer and

Suich 1980) and use graphical tools to examine the solutions to different weight values (Lu, Anderson-Cook, and Robinson 2011). Two of the commonly used DFs are the additive and multiplicative forms. Using the two-criteria as an example, the additive DF is given by $DF_{\text{add}} = wd_1 + (1-w)d_2$, where d_j denotes the desirability value (scaled to $[0,1]$) for criterion $j = 1, 2$ and w represents the weight for criterion 1. The multiplicative DF is expressed as $DF_{\text{multi}} = d_1^w d_2^{1-w}$. The choice between these DF forms depends on how severely one wants to penalize poor performance of a criterion. In general, the additive DF allows the excellent performance of one criterion to offset the poor performance of another, while the multiplicative DF places a more severe penalty on poor performance of any criterion.

Here we illustrate the method using the additive DF. Figure 3b shows the selected designs (shown as the blue points) from the thinned PF of 20 designs (in Figure 2b). By examining 1000 weight combinations

(multiples of 0.001 over the range of [0,1]) for the weight for $G_{\text{rel-eff}}$, the method selects 12 designs which are optimal for at least one set of explored weight choice. The remaining 8 designs are never optimal for any weight choice under the chosen DF form and hence are discarded from further consideration.

By using the DF, we reduce the solutions from 20 on the thinned PF to 12 that are optimal for some weight combinations. To further examine the performance of the solutions and how they align with user priorities (represented by the different weights), Figure 4 shows several graphical tools for examining the trade-offs of the solutions and their robustness to the subjective weight choices. Figure 4a shows the trade-off plot of the criterion values for all 12 solutions from Figure 3b sorted from low to high $G_{\text{rel-eff}}$. The criterion values are plotted on the desirability scale in [0,1] with their original I_{eff} and $G_{\text{rel-eff}}$ values shown on the extended side axes. In this plot, we see that the solutions on the far left and right represent designs with near optimal performance on one criterion but poor performance on the other. Designs in the middle of the plot represent more balanced solutions with good performance on both criteria.

The mixture plot in Figure 4b shows the best designs for different weight combinations. The bottom and top axes show the additive DF weight values for $G_{\text{rel-eff}}$ and I_{eff} and the line segments in the middle show the range of weights for which each of the 12 designs are optimal. For example, we can see the design 1 is I -optimal and design 10501 is G -optimal, respectively. Design 5502 is optimal when I_{eff} and $G_{\text{rel-eff}}$ are equally important ($w = 0.5$). The length of each line segment indicates the robustness of that solution to different weight choices. For example, design 3307 is optimal for a large range of weights between 0.62 and 0.8 for I_{eff} , while design 6124 is optimal when w for $G_{\text{rel-eff}}$ is between 0.64 and 0.78. This summary is useful when there is disagreement or uncertainty about the desired w value, or to reach consensus among different user priorities.

The last graphical summary shown in Figure 4c is the synthesized efficiency plot (Lu and Anderson-Cook 2012), which shows the expected performance of each individual solution compared to the best possible design for different w values. In the plot, each line segment represents the synthesized efficiency (i.e., the relative performance compared to the best possible at a particular weight) of a design over the range of all possible weights. The white-gray-black shades represent the highest to the lowest efficiency values with each change in shade representing a drop of 5% in

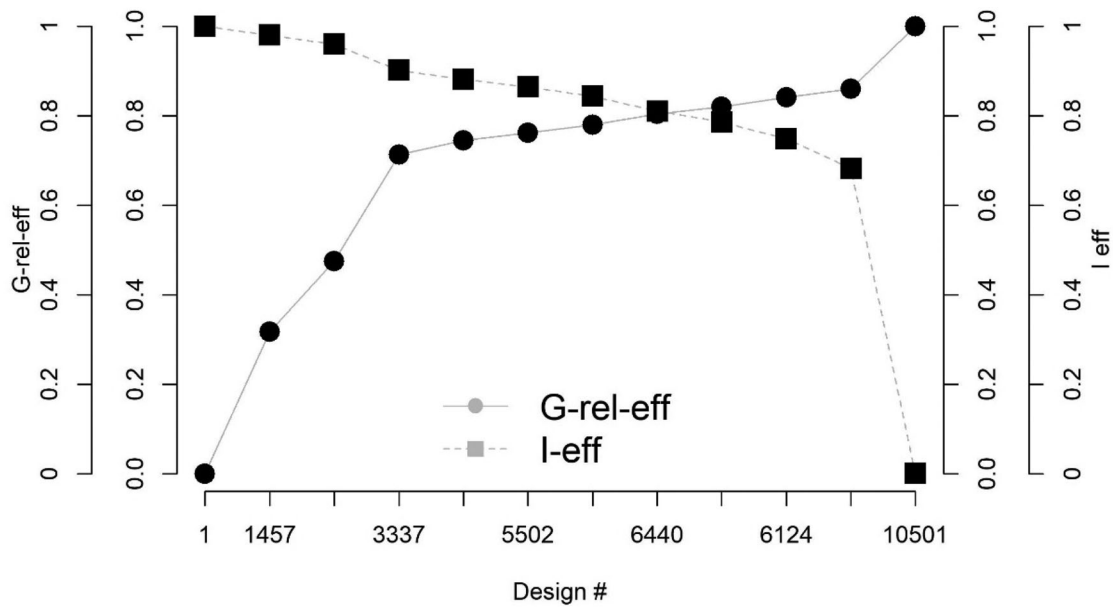
efficiency. For example, design 5502, which is optimal for $w = 0.5$ is at least 95% efficient (white) for w from 0.2 to close to 0.7 for $G_{\text{rel-eff}}$. This indicates great robustness of performance over a wide range of weight values. So, if our range of interesting weights fall into this region, then design 5502 offers excellent performance. Note that design 5502 is also shown in Figures 1a and 3a as the orange point on the PF and its FDS curve is shown in Figure 1b. If one criterion (say I_{eff}) was considered more important than the other, then design 2500 (red point in Figure 1a and red FDS curve in Figure 1b) might be preferred. The combination of this set of graphical summaries allows us to understand and compare individual solutions, and judge how they match desired user priorities, along with their robustness. This facilitates discussion (Anderson-Cook and Lu 2018) and informs decision-making during design selection.

3.3. Optional final step: Fine tuning for the final choice

Once we have selected the best solution from the thinned PF using one of the methods above, it can be helpful to examine solutions from the original rich PF around the selected one to see if any fine tuning of the solution is desired. We could examine the actual design structure, the FDS plot and the predicted variance profile throughout the input space to choose the final winner. Another aspect that should be considered is the granularity of design factor levels that can be implemented in a real experimental setting. Since the selected designs are from a continuous input space, rounding the factor levels to implementable levels might lead to minor changes in the design performance. Examining the performance of the actual design(s) that can be implemented around the chosen candidate could help us select a realistic best design.

4. G- and I-efficient Pareto fronts for $K = 2$, $N = 6$ through 12

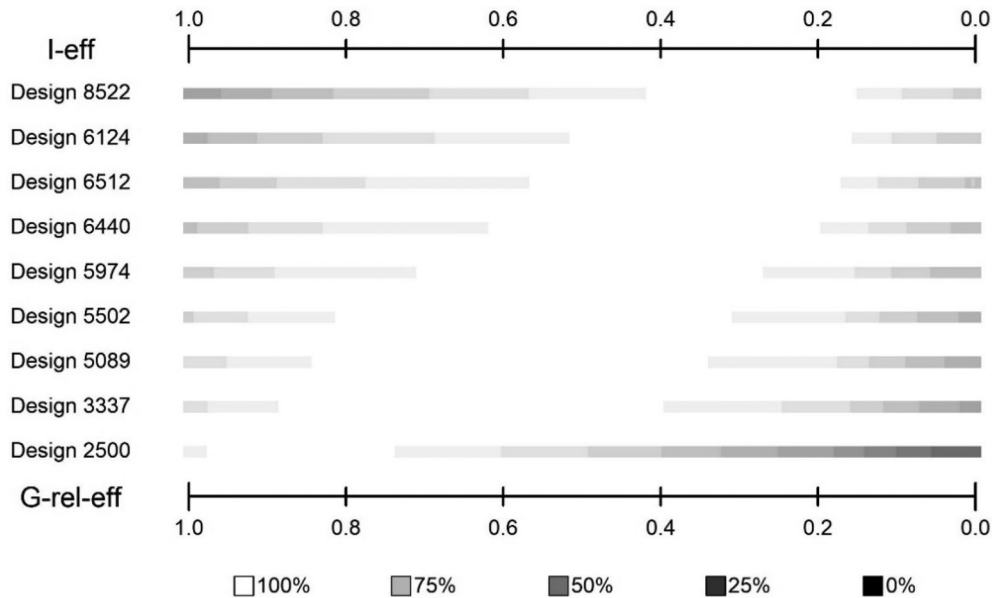
In the previous section, we described approaches to select a best design to match the goals of the experiment for the case $K = 2$, $N = 9$. This same process can be used for any design scenario where a PF is available with an identified suite of non-dominated solutions for the chosen criteria of interest. In this section, we provide the PFs for commonly used design sizes for estimating the second order response surface model for $K = 2$. We complement that with FDS plots



(a) Trade-off plot



(b) Mixture plot



(c) Synthesized efficiency plot

Figure 4. Graphical tools to support further selection and comparison of promising designs from the thinned PF. (a) The tradeoff plot of selected solutions based on the additive desirability function over a variety of possible weight choices; (b) the mixture plot of optimal designs over a fine grid of all possible weights; (c) The synthesized efficiency plot of most promising designs compared with the best possible over the entire range of possible weights.

of five highlighted designs which span the breadth of the PF for each scenario.

First, we consider the PFs shown in Figure 5 for the $K=2$ cases with design sizes of $N=6, 7, 8, 10, 11, 12$. These complement the PF summaries shown in Figure 1a for $K=2$ and $N=9$, and are common design sizes for fitting a second-order response surface model and match the design sizes shown in Borkowski (2003). Table 1 shows some summaries of the different PFs and highlights that the degree of tradeoff between G - and I -efficiency varies considerably depending on the chosen size of the experiment. From Table 1, we see that for the small design sizes ($N=6, 7$ and 8), there is a quite severe cost for pursuing the I -optimal design without any consideration for the $G_{\text{rel-eff}}$. In each case the worst-case RPV is considerably inflated with the I -optimal design, with it being $1/0.563 = 1.776$ to $1/0.650 = 1.538$ times larger than for the G -optimal design. For the larger design sizes ($N=9$ to 12) the penalty for focusing just on I -optimality is not as severe, but still leads to larger worst-case RPV that are $1/0.860 = 1.163$ to $1/0.720 = 1.389$ times larger than the G -optimal design.

When we consider the degree of tradeoff from looking at the I_{eff} of the G -optimal designs we see that there is considerable variation in performance. For the $N=6$ case, we are able to achieve over 92% I_{eff} , or an average RPV that is just $1/0.923 = 1.083$ times bigger than the best possible for the I -optimal design. Depending on the design size, the degree of sacrifice for I_{eff} ranges from 71.4% (with the average RPV 1.4 times larger than the I -optimal design) to 92.3%. The degree of tradeoff does not change monotonically across design sizes, and so considering each scenario separately to understand what is possible is important. Some design size scenarios have nearly equal tradeoffs between I_{eff} and $G_{\text{rel-eff}}$ ($N=12$), while others have substantial differences in the degree of sacrifice needed to optimize one criterion. Finally, the last column of Table 1 shows whether the PF includes a design that is at least 90% efficient for both I_{eff} and $G_{\text{rel-eff}}$ simultaneously. These designs might be promising for easy team consensus.

Next, we consider the shapes of the PFs for the different cases. In the example shown in the Introduction for $K=2$ and $N=9$, the PF was convex up and curved toward the Utopia point. This is the ideal shape for the PF as it suggests that it is possible to find a design that performs well for both the criteria simultaneously. Across the new scenarios shown in Figure 5, we see that the shapes of the PFs vary considerably, with $N=6, 7, 8, 12$ showing some degree of

curvature toward the UP. The case with $N=10$ shows the most severe tradeoff between the two criteria as it does not have intermediate points on the PF for designs close to the UP. Some scenarios have discontinuities, where one of the criteria drops sharply in value as we improve the other criterion. The case with $N=11$ shows a dramatic discontinuity as we move to further improve I_{eff} from $G_{\text{rel-eff}}$ close to 95%.

To fully understand what these tradeoffs mean in practice, we now examine the FDS plots for the five highlighted designs on each PF. The I -optimal (blue) and G -optimal (green) designs represent the extremes of what is achievable when we focus exclusively on one criterion, while the other three designs (red, purple, and orange) show what is possible when both criteria are considered and a suitable balance in performance are sought. Figure 6 shows the complementary plot to Figure 1b for the design scenarios summarized in Figure 5.

Recall that for the FDS plots (Zahran, Anderson-Cook, and Myers 2003), the effect of changing the G -efficiency is observed at the far-right side of the plot, where the worst-case RPV is shown. Changes in the I_{eff} are captured with the average height of the curve across the entire left-to-right range. We see varied patterns in the differences between the five FDS curves for each scenario. For $N=6$, the curves look very similar for most of the range, with separation only occurring to the extreme right of the plot for the worst-case. For many of the other scenarios, the range of the five lines at different fractions of the design space can be substantial, highlighting that there are strong differences in performance throughout the design space depending on which design is selected. For most of the plots, the green G -optimal design rises steadily in near linear fashion from early on the left side of the plot but then has a more moderate worst-case variance. Alternately, the I -optimal design stays flatter longer on the left side of the plot, with a larger proportion of the design space with good prediction performance. However, this comes with the dramatic upturn in the curve at the right side of the plot, with a small fraction having substantially worse RPV.

The three designs from the middle of the PF represent opportunities to mitigate the worst features of either the G - or I -optimal designs. In general, they have larger ranges of better prediction performance (lower values from the left through the middle of the FDS proportions) than the G -optimal design, and less severe worst-case prediction performance (smaller maximum values) than the I -optimal design. Depending on the experimenter's preferences, how to emphasize these two aspects of the RPV will vary, but the five FDS curves

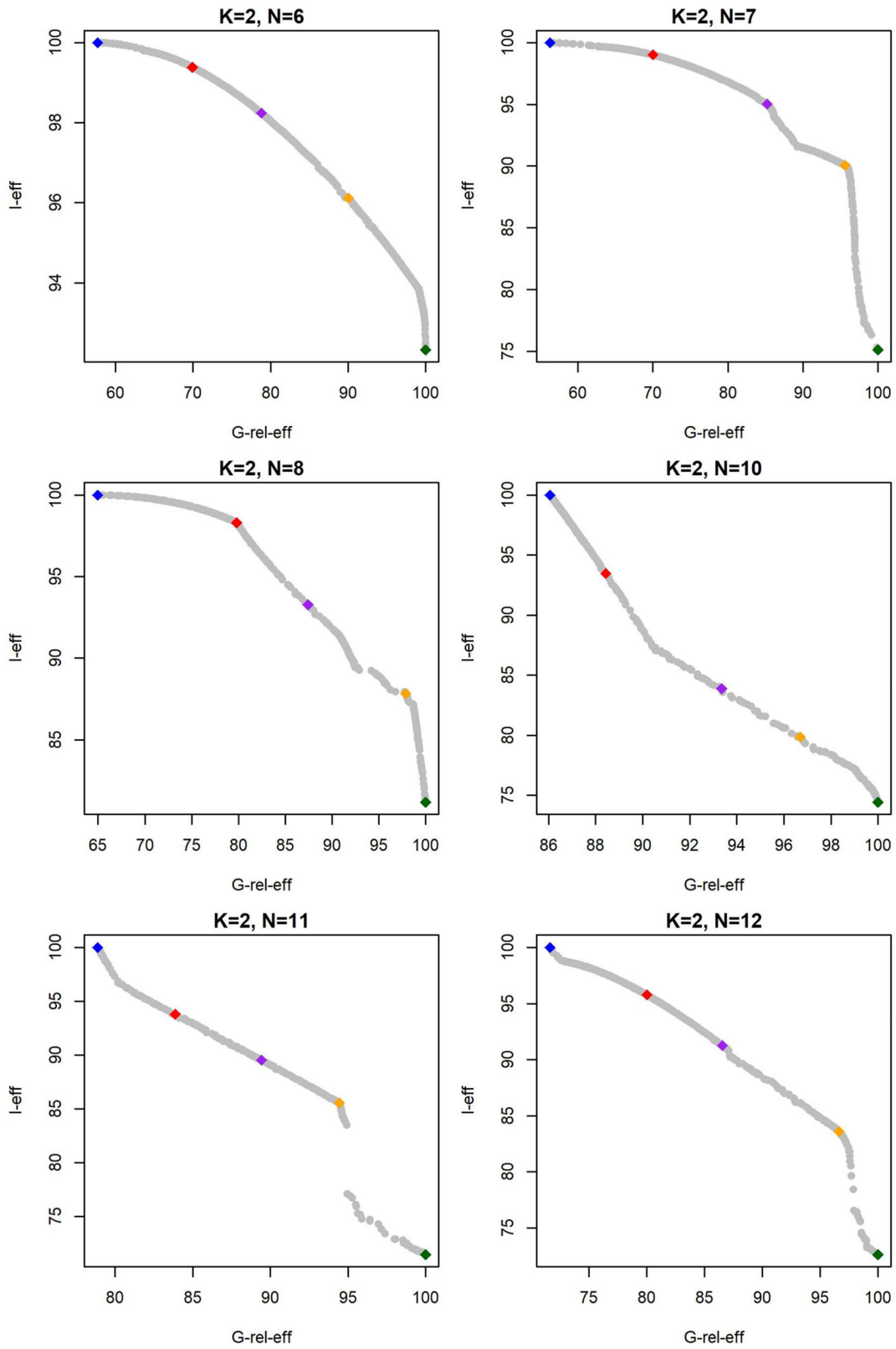


Figure 5. Pareto fronts for the $K=2$ cases with design sizes of $N=6$ through 12 to complement Figure 1a. The five highlighted design solutions represent the I -optimal (blue) and G -optimal (green) designs, as well as three promising solutions (red, purple and orange) from the thinned Pareto front.

Table 1. Summary of Pareto Fronts for $K=2$ and $N=6$ to 12 with the spreadsheets provided in SM.

Design Size	Number of Designs on PF	$G_{\text{rel-eff}}$ of I -optimal design	I_{eff} of G -optimal design	Existence of 90+% $G_{\text{rel-eff}}$ & 90+% I_{eff}
6	2334	57.7	92.3	Y
7	1590	56.3	75.1	Y
8	2107	65.0	81.2	N
9	2508	70.2	80.2	Y
10	605	86.1	74.4	N
11	580	78.9	71.4	N
12	950	71.6	72.6	N

show a diverse set of alternatives for each scenario from which to understand basic patterns.

Table 2 shows details about the 5 designs selected for each scenario with the I_{eff} and $G_{\text{rel-eff}}$ values for each choice. The spreadsheets for $K=2$ scenarios in the SM contain complete information for all the PF designs.

Another aspect to consider is the absolute magnitude of the RPV between scenarios. The summarizes in the PFs of Figures 1a and 5 are scaled to show what is possible relative to the best attainable within that design size, with 100% representing either the G - or the I -optimal design performance. However, with the FDS plots in Figures 1b and 6, we can make direct comparisons between design sizes. The vertical axis shows the RPV (defined in Section 2.1), which is comparable between scenarios. The vertical axes for each of the plots fills the range of values for that scenario, but to compare between design sizes, actual RPV values should be noted.

When we compare performance across the design sizes, we see that increasing the design size from $N=6$ to $N=9$ and then up to $N=12$, yields an improvement in the worst-case RPV from 1.334 to 0.792 to 0.567 times the natural variation (these are largest RPVs of the G -optimal designs for each N). This disproportionate improvement shows how the extreme values of the prediction performance improve with more experimental runs. When we examine the average RPVs for the I -optimal designs, we see that RPV improves from 0.763 to 0.427 to 0.304 for the same three scenarios, showing improvements in the average RPV as sample size increases.

Across the different design sizes for $K=2$, we see different degrees of tradeoff between G - and I -efficiency, differently shaped PFs and substantial changes in the absolute prediction variance magnitude.

5. G - and I -efficient Pareto fronts for $K=3$, $N=10$ through 16 and $K=4$, $N=15$, 17 and 20

In the previous section, we considered designs for two continuous factors. In this section we explore designs

that include three or four continuous input factors. All of the designs on each of the PFs for all cases described are provided in the repository in the SM.

First, consider $K=3$. With an intercept, 3 main effects, 3 quadratic terms and 3 two-factor interactions, the $N=10$ design is saturated. We consider larger designs ($N=11$ to 16) that allow for better estimation of the model terms, and hence also improved prediction variance throughout the input space. For brevity, we only show plots of the even sized designs ($N=10$, 12, 14 and 16) in the main paper, with the odd sized cases ($N=11$, 13, 15) presented in Part 3 of the SM. Table 3 shows comparable summaries for all 7 scenarios (as in Table 1 for the 2-factor case) and highlights that the degree of tradeoff between G - and I -efficiency varies considerably depending on the chosen size of the experiment. Detailed discussion about differences in performance and alternative solutions are provided in Part 4 of the SM.

Similar to the $K=2$ cases, the shapes of the PFs for the different cases for $K=3$ show considerable differences across the range of scenarios (shown in Figure 7 for the even sized cases, and in the SM for the odd sized cases). The ideal shape for the PF of convex up and curved toward the Utopia point occurs for several cases ($N=12$ and 13). This suggests that it is possible to find a design from the interior of the PF that performs well for both criteria simultaneously. The case with $N=10$ shows the most severe tradeoff between the two criteria as the PF curves away from the UP with few intermediate points on the PF for designs which balance the two criteria well. For $N=11$, the PF is relatively straight with nearly proportional trade-off between criteria across the range of designs. Several scenarios have substantial discontinuities ($N=10$, 14, 15, 16), where one criterion drops sharply as we improve the other criterion (Table 4).

To understand more fully what these tradeoffs mean for the distribution of the RPV, we consider the FDS plots in Figure 8 (for the even sized designs) and the SM (for the odd sized designs) for the five highlighted designs on each PF in Figure 7. The I -optimal (blue) and G -optimal (green) designs represent the

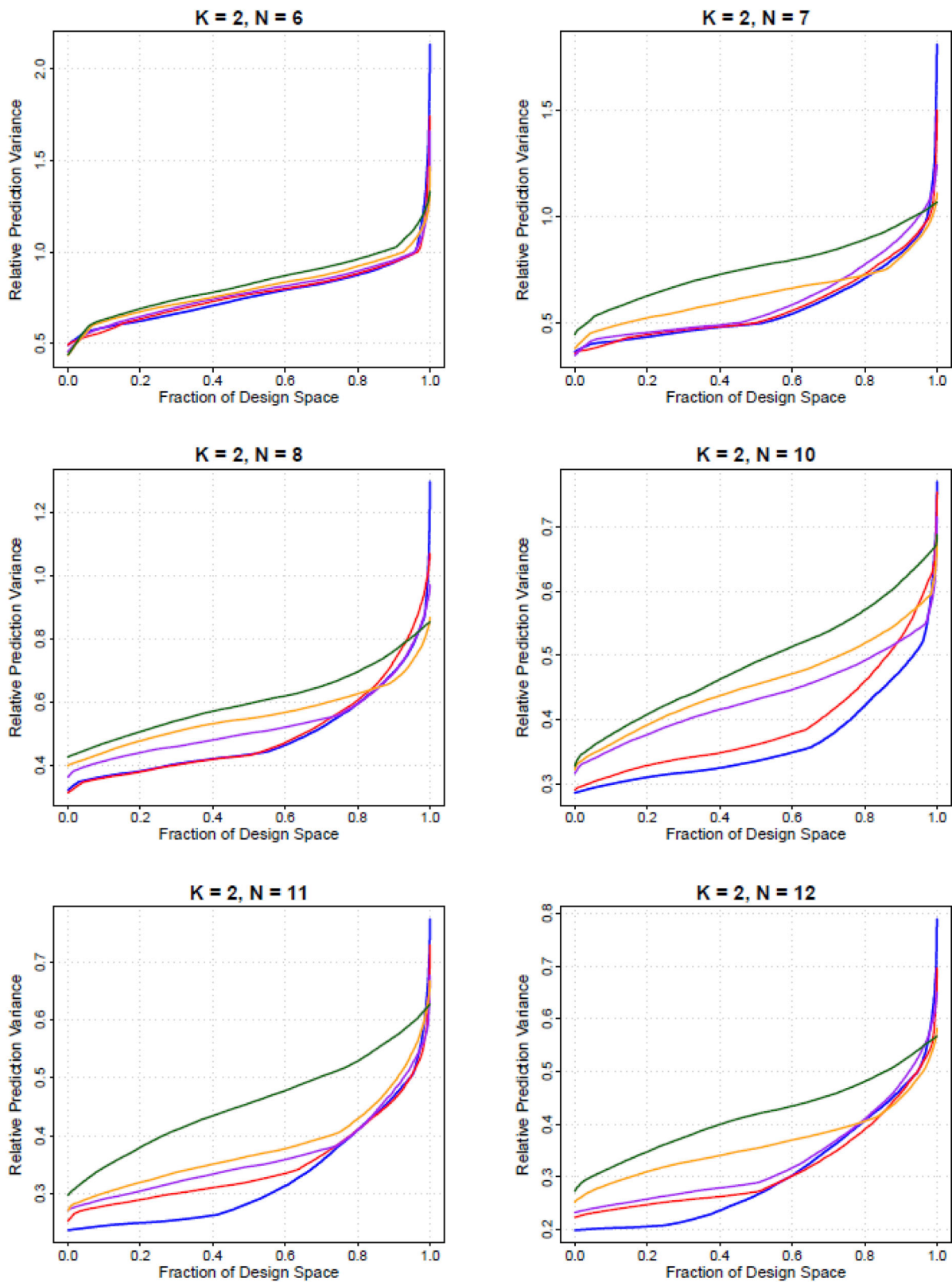


Figure 6. FDS plots for the $K=2$ cases with design sizes of $N=6$ through 12 to complement Figure 1b. The five curves represent the l -optimal (blue) and G -optimal (green) designs, as well as three promising solutions (red, purple and orange) from the thinned Pareto front.

Table 2. Summary of performance for five selected designs in Figures 5 and 6 for $K=2$ and $N=6$ to 12.

Design size	I-optimal (Blue)	Red	Purple	Orange	G-optimal (Green)
6	(100,57.7)	(99.4,69.9)	(98.2,78.9)	(96.1,90.3)	(92.3,100)
7	(100,56.3)	(99.0,70.0)	(95.0,85.3)	(90.1,95.6)	(75.1,100)
8	(100,65.0)	(98.3,79.8)	(93.3,87.4)	(87.8,97.9)	(81.2,100)
9	(100,70.2)	(99.2,84.4)	(97.3,92.9)	(95.0,95.3)	(80.2,100)
10	(100,86.1)	(93.5,88.4)	(83.9,93.4)	(79.9,96.7)	(74.4,100)
11	(100,78.9)	(93.8,83.9)	(89.5,89.4)	(85.5,94.4)	(71.4,100)
12	(100,71.6)	(95.8,80.0)	(91.3,86.5)	(83.6,96.6)	(72.6,100)

The results shown are (I -efficiency, relative G -efficiency).

Table 3. Summary of Pareto Fronts for $K=3$ and $N=10$ to 16.

Design Size	Number of Designs on PF	$G_{\text{rel-eff}}$ of I -optimal design	I_{eff} of G -optimal design	Existence of 90+% $G_{\text{rel-eff}}$ & 90+% I_{eff}
10	258	68.0	81.3	N
11	319	60.8	69.3	N
12	197	57.6	73.6	N
13	155	52.4	72.0	N
14	511	58.9	92.0	Y
15	160	96.8	67.3	Y
16	246	90.8	67.2	Y

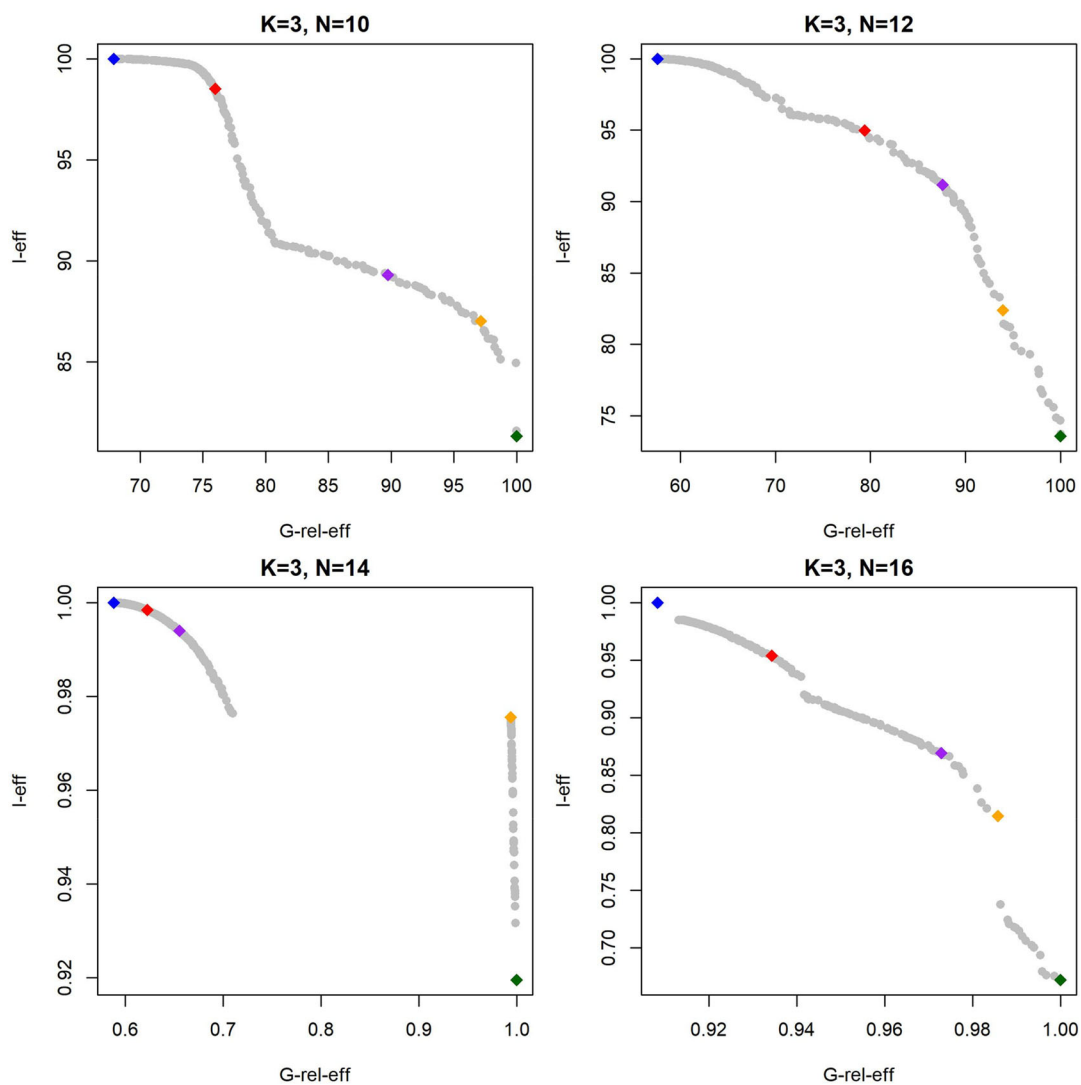


Figure 7. Pareto fronts for the $K=3$ cases with design sizes of $N=10, 12, 14,$ and 16 . The five highlighted design solutions represent the I -optimal (blue) and G -optimal (green) designs, as well as three promising solutions (red, purple and orange) from the thinned Pareto front.

extremes of the distributions of RPV if we focus exclusively on one criterion. The other three designs (red, purple and orange) show possibilities when a balance between the two criteria is sought.

Across the design sizes for $K=3$, we see a variety of degrees of tradeoff between G - and I -efficiency, different

Table 4. Summary of performance for five selected designs in Figures 7 and 8 for $K=3$ and $N=10$ to 16.

Design size	I -optimal (Blue)	Red	Purple	Orange	G -optimal (Green)
10	(100,67.9)	(98.5,76.0)	(89.3,89.7)	(87.0,97.2)	(81.3,100)
11	(100,60.8)	(92.5,78.7)	(86.6,86.7)	(77.8,97.0)	(69.3,100)
12	(100,57.6)	(95.0,79.4)	(91.2,87.6)	(82.4,94.0)	(73.6,100)
13	(100,52.4)	(98.6,72.8)	(92.0,89.0)	(82.9,97.1)	(72.0,100)
14	(100,58.9)	(99.8,62.3)	(99.4,65.6)	(97.6,99.5)	(92.0,100)
15	(100,96.8)	(94.1,98.1)	(91.7,99.1)	(88.6,99.9)	(67.3,100)
16	(100,90.8)	(95.4,93.4)	(86.9,97.3)	(81.5,98.6)	(67.2,100)

The results shown are (I -efficiency, relative G -efficiency).

shapes of PFs and improvements in the absolute relative prediction variance magnitude as the design gets larger. Next, consider $K=4$. With the full second-order model including an intercept, 4 main effects, 4 quadratic terms and 6 two-factor interactions, the $N=15$ design is saturated. The larger $N=17$ and 20 cases allow for better estimation of the model terms, and hence also improved prediction variance throughout the input space. Table 5 shows PF summaries for the three design scenarios and highlights the degree of tradeoff between G - and I -efficiency. Figure 9 shows the PFs for each of the cases and illustrates the degree of tradeoff required to balance the two objectives. Note that the number of solutions on the PF has again diminished relative to the smaller dimension cases. The curse of dimensionality becomes stronger for the $K=4$ larger design cases as the PSO search algorithm has a dramatically harder task to thoroughly explore the

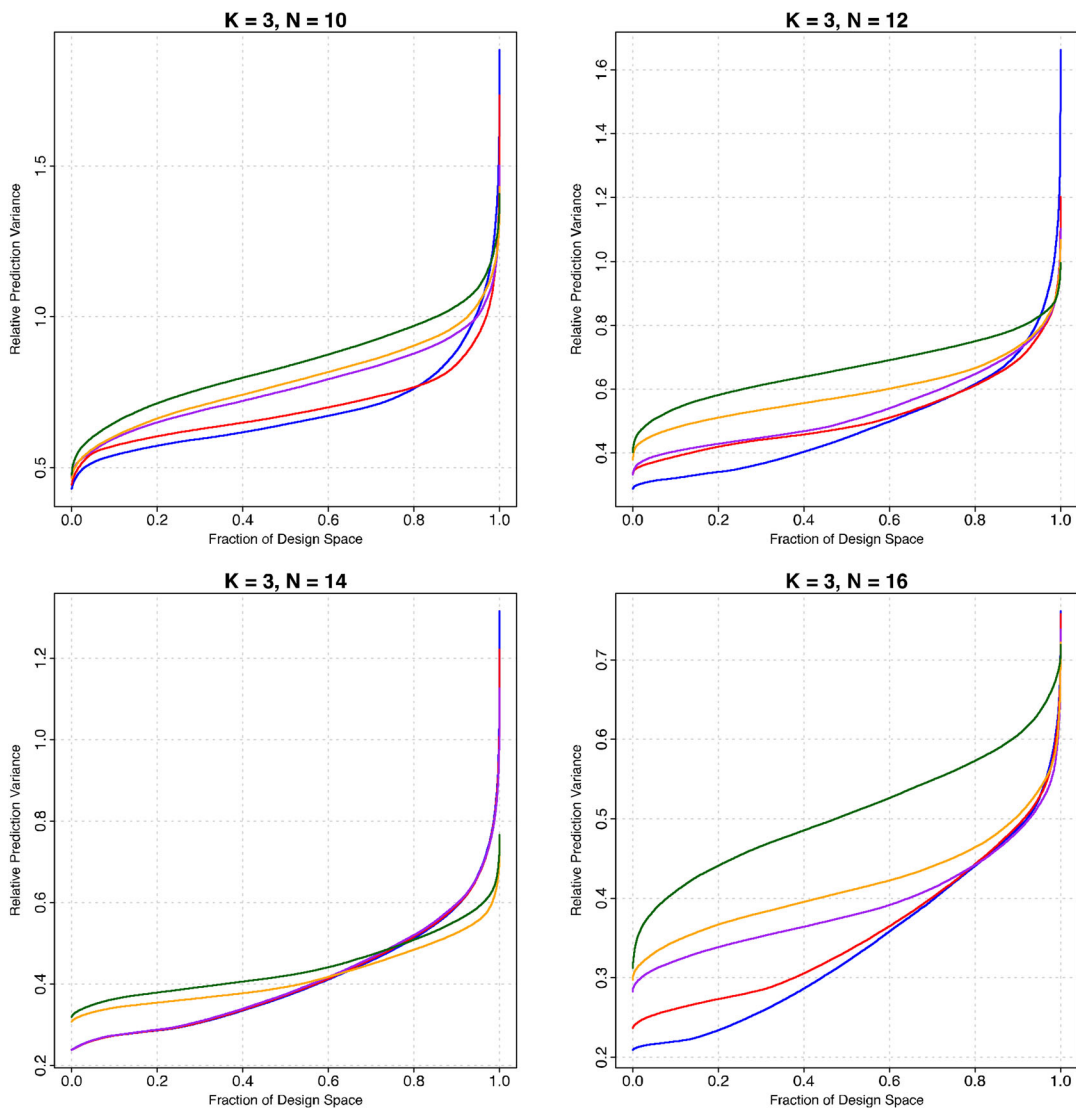


Figure 8. FDS plots for the $K=3$ cases with design sizes of $N=10, 12, 14$ and 16 to complement Figure 7. The five curves represent the I -optimal (blue) and G -optimal (green) designs, as well as three promising solutions (red, purple and orange) from the thinned Pareto front.

Table 5. Summary of Pareto Fronts for $K=4$ and $N=15, 17$, and 20 .

Design Size	Number of Designs on PF	$G_{\text{rel-eff}}$ of I -optimal design	I_{eff} of G -optimal design	Existence of 90+% $G_{\text{rel-eff}}$ & 90+% I_{eff}
15	492	52.1	69.4	N
17	201	69.8	61.1	Y
20	120	64.4	68.4	N

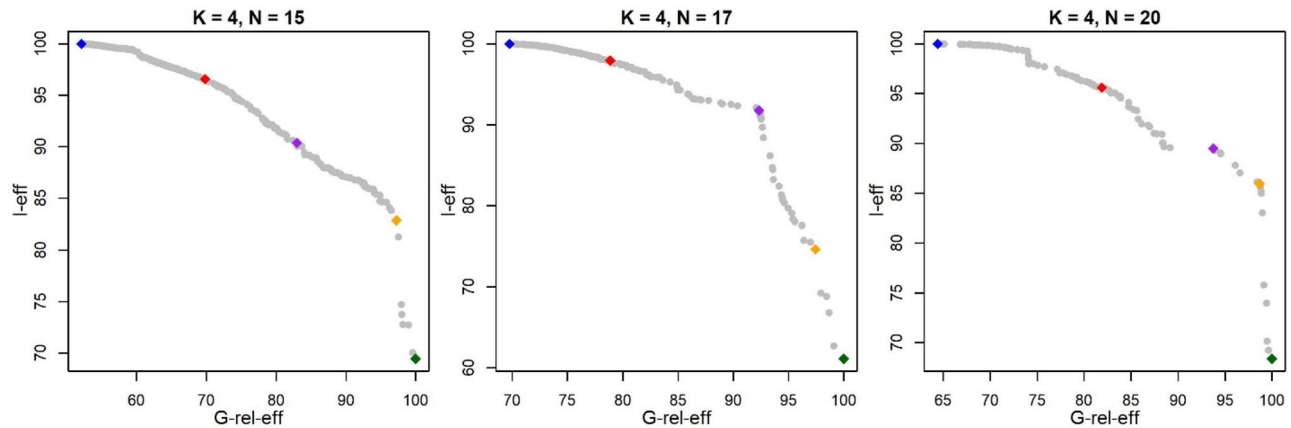
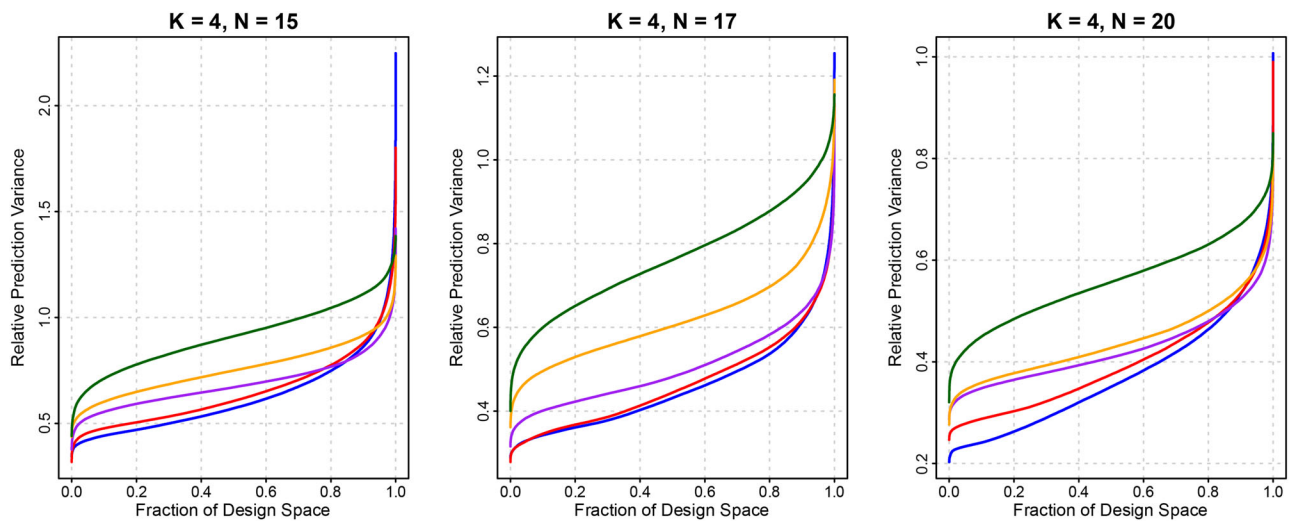

Figure 9. Pareto fronts for the $K=4$ cases with design sizes of $N=15, 17$ & 20 . The five highlighted design solutions represent the I -optimal (blue) and G -optimal (green) designs, as well as three promising solutions (red, purple and orange) from the thinned Pareto front.

Figure 10. FDS plots for the $K=4$ cases with design sizes of $N=15, 17$ and 20 to complement [Figure 9](#). The five curves represent the I -optimal (blue) and G -optimal (green) designs, as well as three promising solutions (red, purple and orange) from the thinned Pareto front.

Table 6. Summary of performance for five selected designs in [Figures 9](#) and [10](#) for $K=4$ and $N=15, 17$, and 20 .

Design size	I -optimal (Blue)	Red	Purple	Orange	G -optimal (Green)
15	(100,52.1)	(96.6,69.9)	(90.4,83.0)	(82.9,97.2)	(69.4,100)
17	(100,69.8)	(97.9,78.9)	(91.7,92.3)	(74.7,97.4)	(61.1,100)
20	(100,64.4)	(95.6,81.9)	(89.5,93.8)	(85.9,98.7)	(68.4,100)

The results shown are (I -efficiency, relative G -efficiency).

possibilities. This results in Pareto fronts that are less rich. However, there are still numerous choices from which to select a tailored design. Detailed discussion

about differences in performance and alternative solutions are provided in Part 5 of the SM.

The FDS plots for the five highlighted designs on each PF allow us to more fully understand what these tradeoffs mean for the distribution of the RPV. In [Figure 10](#), the FDS plots are shown for the five designs highlighted in each scenario of [Figure 9](#). The extremes of the RPV distribution if we focus only on one criterion are denoted by the blue (I -optimal) and green (G -optimal) curves. The other three designs (red, purple, and orange) show some of the compromises that are possible ([Table 6](#)).

6. Conclusions

In this article, we provide suites of designs that allow experimenters to understand the tradeoff between focusing on minimizing the average or worst-case prediction variance performance of designs of different sizes for 2, 3 and 4 continuous input factors. Historically, experimenters were presented with the choice of selecting one of G - or I -optimality, and provided with little information about how the specified design performs for the other metric relative to what was possible. Based on the results from these different scenarios, we have demonstrated that if we focus narrowly on only one of G - or I -optimality, we often get quite poor performance on the other criterion. But if we prioritize good, but not optimal, performance for both criteria it is often possible to find a design that performs well. The Pareto front approach allows each experimenter to see numerous available alternatives and find the right balance between the two objectives for their study.

The incorporation of the particle swarm optimization algorithm for constructing Pareto fronts allows for the exploration of continuous design spaces and the efficient population of these PFs with ample choices across the range of emphases of the two objectives. We acknowledge that the algorithm that generated the provided solutions has the potential to be refined to further improve efficiency and the density of choices available, particularly as the size and dimension of the designs increase. This is a priority for future research.

We selected the 17 scenarios (7 for $K=2$, 7 for $K=3$ and 3 for $K=4$) to show that the degree of tradeoff, the severity of consequence for focusing on a single criterion and the shape of the PF are dependent on design size. For 9 of the 17 scenarios (5 for $K=2$, 3 for $K=3$, 1 for $K=4$), it is possible to identify at least one design with at least 90% efficiency for both criteria. With just a small compromise in one of the criteria away from the optimal design, it is often possible to realize a disproportionate improvement in the other criterion.

With the methodology outlined for selecting between the alternatives, the experimenter has flexibility to match the design to their priorities. As we demonstrated via the examples, the degree of tradeoff between the two objectives varies for different scenarios. Hence it is important to see the options for the particular designed experiment of interest in order to make the best decision. This article offers a structured process for making such a decision based on finding the Pareto set of solutions, reducing choices, and

selecting the final best match with experimental goals. We hope that the sets of designs provided for all the scenarios in the SM facilitate the practical application of this approach and allow for more robust designs to be utilized that have good prediction variance performance across the majority of the design space (with a good average and high I -efficiency) as well as at the edges (where the worst-case often occurs).

The methodology outlined is flexible and can be easily adapted to other situations. For example, if the experimenter wishes to focus on estimation of the model parameters, then D - or A -optimality could be used and the process of constructing the Pareto fronts and then selecting a best design could be easily implemented. Changes to the PSO search algorithm when using any pair of optimality criteria is straightforward. Similarly, if the experimenter wished to consider more than two criteria (say, G -, I - and D -optimality), the method could be easily adapted. The PSO search algorithm would need to be expanded to include a set of weights to cover the space of desired weights (Lu, Anderson-Cook, and Lin 2014) combining all of the criteria. After construction of the PF, the graphical tools would also need to be adapted (Lu, Chapman, and Anderson-Cook 2017) for the larger space of solutions. We caution against using too many criteria simultaneously as balancing multiple objectives can often lead to mediocre performance on several of them. Finally, the methodology could also be expanded beyond designs with all continuous factors to consider those scenarios with categorical design factors, see for example Chen et al. (2014). In this case, the PSO search would need to be adapted to accommodate this new design region. Once the PF is constructed, the remainder of the selection process would match what has been illustrated in this article.

By acknowledging that not all experimental situations require the same solution, we hope to empower experimenters to be easily able to compare alternatives and then choose the design that is the best fit for their priorities.

About the authors

Dr. **Stephen J. Walsh** is a faculty member in the Department of Mathematics and Statistics at Utah State University. He has over a decade of experience practicing as a quality assurance and experimental design expert in laboratory environments. From 2007 to 2011 he was a researcher at Pacific Northwest National Laboratory in Richland, WA. From 2011 to 2017 he worked as a government Statistician and Quality Manager at the International Atomic Energy Agency in Vienna, Austria. His Ph.D., earned in 2021, provided an adaptation of the Particle

Swarm Optimization to solving several difficult optimal design problems. His current research program focuses on the synergistic research bridge between design of experiments and machine learning algorithms.

Lu Lu is an Associate Professor of Statistics in the Department of Mathematics and Statistics at the University of South Florida in Tampa. She was a postdoctoral research associate in the Statistics Sciences Group at Los Alamos National Laboratory. Her research areas include statistical engineering, reliability analysis, design of experiments, response surface methodology, survey sampling, and multiple objective/response optimization.

Christine M. Anderson-Cook is a statistician and Retired Guest Scientist in the Statistical Sciences Group at Los Alamos National Laboratory. Her research areas include statistical engineering, reliability, design of experiments, multiple criterion optimization, and response surface methodology. She is a Fellow of the American Statistical Association (ASA) and the American Society for Quality (ASQ).

Acknowledgments

We first thank Quality Engineering for an expedient, professional, and quality review. We also thank the two anonymous referees whose comments led to substantial improvements to the manuscript.

ORCID

Stephen J. Walsh  <http://orcid.org/0000-0002-0505-648X>

Lu Lu  <http://orcid.org/0000-0001-7754-9374>

Christine M. Anderson-Cook  <http://orcid.org/0000-0002-0165-5565>

References

- Anderson-Cook, C. M., and L. Lu. 2018. "Graphics to facilitate informative discussion and team decision-making" (with discussions). *Applied Stochastic Models in Business and Industry* 34 (6):963–80. doi:10.1002/asmb.2325.
- Borkowski, J. 2003. *Using a genetic algorithm to generate small exact response surface designs*. Saskatchewan, CA: University of Regina.
- Cao, Y., B. J. Smucker, and T. J. Robinson. 2017. A hybrid elitist pareto-based coordinate exchange algorithm for constructing multi-criteria optimal experimental designs. *Statistics and Computing* 27 (2):423–37. doi:10.1007/s11222-016-9630-9.
- Chapman, J. L., L. Lu, and C. M. Anderson-Cook. 2018. Using multiple criteria optimization and two-stage genetic algorithms to select a population management strategy with optimized reliability. *Complexity* 2018:1–18. doi:10.1155/2018/7242105.
- Chen, R. B., S. P. Chang, W. Wang, H.-C. Tung, and W. K. Wong. 2015. Minimax optimal designs via particle swarm optimization methods. *Statistics and Computing* 25 (5): 975–88. doi:10.1007/s11222-014-9466-0.
- Chen, P. Y., R. B. Chen, and W. K. Wong. 2022. Particle swarm optimization for searching efficient experimental designs: A review. *WIREs Computational Statistics* 14 (5): 4301–17. doi:10.1002/wics.1578.
- Chen, R. B., Y. W. Hsu, Y. Hung, and W. Wang. 2014. Discrete particle swarm optimization for constructing uniform design on irregular regions. *Computational Statistics & Data Analysis* 72:282–97. doi:10.1016/j.csda.2013.10.015.
- Chen, R. B., C. H. Li, Y. Hung, and W. Wang. 2019. Optimal noncollapsing space-filling designs for irregular experimental regions. *Journal of Computational and Graphical Statistics* 28 (1):74–91. doi:10.1080/10618600.2018.1482760.
- Chen, Y., and A. Zhou. 2022. Multiobjective portfolio optimization via Pareto front evolution. *Complex & Intelligent Systems* 8 (5):4301–17. doi:10.1007/s40747-022-00715-8.
- Clerc, M. Standard particle swarm optimization. Technical report, HAL Archivesouvertes, 2012.
- Clerc, M., and J. Kennedy. Feb 2002. The particle swarm - explosion, stability, and convergence in a multidimensional complex space. *IEEE Transactions on Evolutionary Computation* 6 (1):58–73. doi:10.1109/4235.985692.
- Deb, K., A. Pratap, S. Agarwal, and T. Meyarivan. 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation* 6 (2): 182–97. doi:10.1109/4235.996017.
- Deb, K., and H. Jain. 2014. An evolutionary many-objective optimization algorithm using reference-point-based non-dominated sorting approach, Part I: Solving problems with box constraints. *IEEE Transactions on Evolutionary Computation* 18 (4):577–601. doi:10.1109/TEVC.2013.2281535.
- Derringer, G., and R. Suich. 1980. Simultaneous optimization of several response variable. *Journal of Quality Technology* 12 (4):214–9. doi:10.1080/00224065.1980.11980968.
- Fox, A. D., Corne, D. W. Mayorga Adame, C. G. Polton, J. A. Henry, L.-A. Roberts, and J. M. 2019. An efficient multi-objective optimization methods for use in the design of marine protected area networks. *Frontiers in Marine Science* 6 (17). doi:https://doi.org/10.3389/fmars.2019.00017.
- Goos, P., and B. Jones. 2011. *Optimal design of experiments: A case study approach*. Hoboken, NJ: John Wiley and Sons, Ltd.
- Gronwald, W., T. Hohm, and D. Hoffmann. 2008. Evolutionary pareto-optimization of stably folding peptides. *BMC Bioinformatics* 9:109. doi:10.1186/1471-2105-9-109.
- Hernandez, L. N., and C. J. Nachtsheim. 2018. Fast computation of exact G-optimal designs via II-optimality. *Technometrics* 60 (3):297–305. doi:10.1080/00401706.2017.1371080.
- Hernández-Díaz, A. G., L. V. Santana-Quintero, C. A. Coello Coello, and J. Molina. 2007. Pareto-adaptive epsilon-dominance. *Evolutionary Computation* 15 (4):493–517. doi:10.1162/evco.2007.15.4.493.
- Jensen, W. A. 2018. Open problems and issues in optimal design. *Quality Engineering* 30 (4):583–93. doi:10.1080/08982112.2018.1517884.
- Laumanns, M., L. Thiele, K. Deb, and E. Zitzler. 2002. Combining convergence and diversity in evolutionary

- multi-objective optimization. *Evolutionary Computation* 10 (3):263–82. doi:10.1162/106365602760234108.
- Lin, C. F. D., C. M. Anderson-Cook, M. S. Hamada, L. M. Moore, and R. R. Sitter. 2015. Using genetic algorithms to design experiments: A review. *Quality and Reliability Engineering International* 31 (2):155–67. doi:10.1002/qre.1591.
- Lu, L., and C. M. Anderson-Cook. 2012. Rethinking the optimal response surface design for a first-order model with two-factor interactions, when protecting against curvature. *Quality Engineering* 24 (3):404–22. doi:10.1080/08982112.2012.629940.
- Lu, L., and C. M. Anderson-Cook. 2013. Adapting the hypervolume quality indicator to quantify trade-offs and search efficiency for multiple criteria decision-making using pareto fronts. *Quality and Reliability Engineering International* 29 (8):1117–33. doi:10.1002/qre.1464.
- Lu, L., and C. M. Anderson-Cook. 2014. Balancing multiple criteria incorporating cost using Pareto front optimization for split-plot designed experiments. *Quality and Reliability Engineering International* 30 (1):37–55. doi:10.1002/qre.1476.
- Lu, L., and C. M. Anderson-Cook. 2021. Input-response space-filling (IRSF) designs. *Quality and Reliability Engineering International* 37 (8):3529–51. doi:10.1002/qre.2931.
- Lu, L., Anderson-Cook, C. M. Robinson, and T. J. 2011. Optimization of designed experiments based on multiple criteria utilizing a Pareto frontier. *Technometrics* 53 (4): 353–65. doi:10.1198/TECH.2011.10087.
- Lu, L., C. Anderson-Cook, and M. Zhang. 2021. Understanding the merits of winning solutions from a data competition for varied sets of objectives. *Statistical Analysis and Data Mining: The ASA Data Science Journal* 14 (6):556–74. doi:10.1002/sam.11494.
- Lu, L., C. M. Anderson-Cook, and D. Lin. 2014. Optimal designed experiments using a pareto front search for focused preference of multiple objectives. *Computational Statistics & Data Analysis* 71:1178–92. doi:10.1016/j.csda.2013.04.008.
- Lu, L., J. L. Chapman, and C. M. Anderson-Cook. 2013. A case study on selecting a best allocation of new data for improving the estimation precision of system and sub-system reliability using Pareto fronts. *Technometrics* 55 (4):473–87. doi:10.1080/00401706.2013.831776.
- Lu, L., J. L. Chapman, and C. M. Anderson-Cook. 2017. Multiple response optimization for higher dimensions in factors and responses. *Quality and Reliability Engineering International* 33 (4):727–44. doi:10.1002/qre.2051.
- Lu, L., M. Li, and C. M. Anderson-Cook. 2016. Multiple objective optimization in reliability demonstration tests. *Journal of Quality Technology* 48 (4):326–42. doi:10.1080/00224065.2016.11918172.
- Lukemire, J., A. Mandal, and W. K. Wong. 2019. d-qps: A quantum-behaved particle swarm technique for finding D-optimal designs with discrete and continuous factors and a binary response. *Technometrics* 61 (1):77–87. doi:10.1080/00401706.2018.1439405.
- Mak, S., and V. R. Joseph. 2018. Minimax and minimax projection designs using clustering. *Journal of Computational and Graphical Statistics* 27 (1):166–78. doi:10.1080/10618600.2017.1302881.
- Marler, R. T., and J. S. Arora. 2004. Survey of multi-objective optimization methods for engineering. *Structural and Multidisciplinary Optimization* 26 (6):369–95. doi:10.1007/s00158-003-0368-6.
- Meyer, R. K., and C. J. Nachtsheim. 1995. The coordinate-exchange algorithm for constructing exact optimal experimental designs. *Technometrics* 37 (1):60–9. doi:10.1080/00401706.1995.10485889.
- Myers, R. H., D. C. Montgomery, and C. M. Anderson-Cook. 2016. *Response surface methodology: Process and product optimization using designed experiments*. 4th ed. New York: Wiley.
- Reuter, H. 1990. An approximation method for the efficiency set of multiobjective programming problems. *Optimization* 21 (6):905–11. doi:10.1080/02331939008843621.
- Rodriguez, M., B. Jones, C. M. Borror, and D. C. Montgomery. 2010. Generating and assessing exact G-optimal designs. *Journal of Quality Technology* 42 (1):3–20. doi:10.1080/00224065.2010.11917803.
- Shi, Y., Z. Zhang, and W. Wong. 2019. Particle swarm-based algorithms for finding locally and Bayesian D-optimal designs. *Journal of Statistical Distributions and Applications* 6 (1):1–17. doi:10.1186/s40488-019-0092-4.
- Trautmann, H., and J. Mehnen. 2009. Preference-based Pareto optimization in certain and noisy environments. *Engineering Optimization* 41 (1):23–38. doi:10.1080/03052150802347926.
- Walsh, S. J. 2021. *Development and applications of particle swarm optimization for constructing optimal experimental designs*. Ph.D. Dissertation., Montana State University. <https://scholarworks.montana.edu/xmlui/handle/1/16309>.
- Walsh, S. J., and J. J. Borkowski. 2022. Generating exact optimal designs via particle swarm optimization: assessing efficacy and efficiency via case study. *Quality Engineering* 1–20. doi:10.1080/08982112.2022.2127364.
- Walsh, S. J., and J. J. Borkowski. 2022. Improved G-optimal designs for small exact response surface scenarios: fast and efficient generation via particle swarm optimization. *Mathematics* 10 (22):4245. doi:10.3390/math10224245.
- Wong, W. K., R. B. Chen, C. C. Huang, and W. Wang. 2015. 06 A modified particle swarm optimization technique for finding optimal designs for mixture models. *PLoS ONE* 10 (6):e0124720. doi:10.1371/journal.pone.0124720.
- Zahrán, A., C. M. Anderson-Cook, and R. H. Myers. 2003. Fraction of design space to assess the prediction capability of response surface designs. *Journal of Quality Technology* 35 (4):377–86. doi:10.1080/00224065.2003.11980235.