A Case Study on Selecting a Best Allocation of New Data for Improving the Estimation Precision of System and Subsystem Reliability Using Pareto Fronts

Lu Lu , Jessica L. Chapman & Christine M. Anderson-cook

A Case Study on Selecting a Best Allocation of New Data for Improving the Estimation Precision of System and Subsystem Reliability Using Pareto Fronts

Lu Lu

Department of Mathematics and Statistics
University of South Florida
Tampa, FL 33620
(*lulu1@usf.edu*)

Jessica L. Chapman

Department of Mathematics
Computer Science and Statistics
St. Lawrence University
Canton, NY 13617
(*jchapman@stlawu.edu*)

Christine M. Anderson-Cook

Statistical Sciences Group
Los Alamos National Laboratory
Los Alamos, NM 87545
(*candcook@lanl.gov*)

This article demonstrates how the Pareto front multiple objective optimization approach can be used to select a best allocation of new data to collect from among many different possible data sources with the goal of maximally reducing the width of the credible intervals of system and two subsystem reliability estimates. The method provides a streamlined decision-making process by identifying a set of noninferior or admissible allocations either from a given set of candidate choices or through a global optimization search and then using graphical methods for selecting the best allocation from the set of contending choices based on the specific goals of the study. The approach allows for an easy assessment of the tradeoffs between criteria and the robustness of different choices to different prioritization of experiment objectives. This is important for decision makers to make a defensible choice of a best allocation that matches their priorities as well as to quantify the anticipated advantages of their choice relative to other options. The method is demonstrated on a small nonaging series system with two subsystems comprised of six components for a total of nine possible data sources. We first consider finding the Pareto front of superior allocations based on 60 logistically viable candidates that have been identified, and second, optimizing over all possible allocations within the allowable fixed budget and comparing how global solutions perform relative to the logistically viable choices. We develop a new search algorithm to populate the Pareto front while taking into account the different costs of the data sources. The method generalizes easily to other system structures and flexible objectives of interest. In addition, a new Fraction of Weight Space plot (FWS) is proposed to provide a simple comparison between different solution choices by summarizing individual performance over the entire weighting space. This article has supplementary materials and computer code available online.

KEY WORDS: Complex system reliability; Genetic algorithm; Multiple data sources; Optimizing multiple objectives; Resource allocation; Sequential data collection.

1. INTRODUCTION

Managers of stockpiles or populations of complex systems often have to make difficult decisions about how to estimate the probability of successful operation, or reliability, of their systems using data from a variety of different sources, such as different levels of a system. As a hypothetical example, consider a relatively simple series system (S) comprised of two subsystems: the mechanical subsystem (M) with two components (M1 and M2) and the electrical subsystem (E) with four components (E1, E2, E3, and E4). If the manager has the ability to collect binary (pass/fail) data at all levels of the system, the manager must determine the best combination of data from these sources to meet his or her needs. Martz, Waller, and Fickas (1988), Martz and Waller (1990), Johnson et al. (2003), and Anderson-Cook

et al. (2007) considered Bayesian analyses of complex systems based on system, subsystem, and component level binary data. Using data from different levels of the system can reduce the need for destructive and expensive full-system tests, and can allow supplementary information to be incorporated into the analysis. Hamada et al. (2008) provided a broad discussion of the topic. This framework of considering multiple data sources in a single analysis to improve estimation while reducing cost is used in a diverse set of applications (Anderson-Cook 2009b).

In this article, we consider the situation where some pass/fail data have already been collected, and there is an opportunity to collect additional data subject to a budget constraint, with the goal of improving the precision of the reliability estimates of the system and subsystems. There are many different possibilities of what new data to collect within a fixed budget from among the multiple sources of data available. This problem, called "resource allocation," is closely related to sequential experimentation and seeks to select a best set of new data based on quantitative measures of the anticipated improvement in precision when the new data are combined with the original and the analysis is updated. Anderson-Cook, Graves, and Hamada (2009) presented a computationally intensive approach for this problem. Chapman, Morris, and Anderson-Cook (2012) suggested a more efficient approach for assessing different allocations when the reliability of the system is not thought to change as a function of its age. In these previous cases, the choice of a best allocation has been based solely on the anticipated improvement in the precision of estimation of system reliability.

However, managers of systems, such as the hypothetical series system, often need to balance several competing objectives in deciding how to best look after the population. Thus, we present new methodology that allows decision makers to simultaneously consider multiple objectives when selecting a best allocation. Here the focus is to improve the precision of estimation for the system and subsystem reliabilities, by comparing alternative allocations based on their anticipated improvements in the credible interval widths for the reliabilities. However, the number of objectives considered, and how to quantify those objectives, can be adapted to the problem at hand. This general approach works for any resource allocation scenario with multiple data sources.

Throughout this article, we present the methods in the context of the hypothetical series system. We refer to the components, subsystems, and system generically as nodes, and we assume that the responses available for collection is pass/fail on all nodes. The cost of the pass/fail data for the different nodes varies, with a full system test being 12 times the cost of component level data for any of the electrical components. The costs per observation for the various nodes, available in Table 1, have been scaled so that the cheapest data sources cost one unit per new observation. While this is a hypothetical example, the key aspects match characteristics of real examples that we have encountered in practice. Note that when we consider the relative cost of the data sources, we find obtaining one observation for all components is cheaper than obtaining one observation for each of the subsystems, which is again cheaper than a single system observation. While common, this will not always be true.

Table 1 displays the data already collected before the start of this resource allocation problem, with the cost of all of the data totaling 325 units. All of the data available at the system, subsystem, and component levels are pass/fail for a single failure mode, and the reliabilities of all of the nodes are assumed to be unchanged as a function of age or time. These assumptions, combined with the relatively small number of nodes, make for a relatively simple version of a complex system. However, the overall approach for multiple criteria optimization for future data collection based on multiple data sources, with potentially different costs, generalizes to more complicated systems with some adaptations.

For a series system, we assume that the system works if and only if all components work appropriately and the components work independently. The reliability, or probability of performing successfully, for any of the parts comprised of multiple components is the product of its corresponding component reliabilities. For the hypothetical system let p_i, $i \in Q = \{M1, M2, E1, E2, E3, E4, M, E, S\}$, denote the reliabilities for the different nodes. The reliability of the mechanical subsystem is $p_M = p_{M1} \cdot p_{M2}$; the reliability of the electrical subsystem is $p_E = p_{E1} \cdot p_{E2} \cdot p_{E3} \cdot p_{E4}$. The reliability of the system can then be expressed in terms of the two subsystems as $p_S = p_M \cdot p_E$.

We model all of the pass/fail data as $x_i \sim \text{Bin}(n_i, p_i)$, where x_i is the number of successful tests out of n_i trials with reliability p_i for $i \in Q$, and let the vectors $\boldsymbol{x}$, $\boldsymbol{n}$, and $\boldsymbol{p}$ denote the aggregate number of successes, trials, and reliability parameter vector, respectively. Further, we assume that the results of the different tests are independent of each other and that a failed higher level test provides no specific information about which component failed. We perform a Bayesian analysis of the existing data using Markov chain Monte Carlo (MCMC) methods to obtain a posterior sample of the node reliabilities. These posterior samples yield the posterior means and credible intervals given in Table 1. The details of the analysis are provided in the supplementary materials I.

Now, consider the scenario where 120 cost units are available to collect additional data, and an overall system manager is tasked with decisions about the overall population of systems.

Table 1. Summary of existing data and information for the hypothetical series system

Part	# Successes	# Trials	Posterior mean (95% Credible interval)	Cost per observation	Cost
M1	11	12	0.927 (0.785, 0.996)	2	24
M2	14	15	0.935 (0.810, 0.997)	2	30
E1	18	20	0.881 (0.724, 0.981)	1	20
E2	23	24	0.947 (0.836, 0.998)	1	24
E3	27	30	0.885 (0.758, 0.970)	1	30
E4	25	27	0.912 (0.785, 0.987)	1	27
Mechanical	7	8	0.866 (0.716, 0.966)	6	48
Electrical	5	10	0.671 (0.527, 0.801)	5	50
System	4	6	0.581 (0.432, 0.722)	12	72
					Total = 325

Hence this manager's interest is in good precision for estimating the system reliability. In addition, there are two subsystem managers responsible for preparing the maintenance resources for each of the subsystems under their control. Their priorities lie in having good precision for estimating the reliability of their subsystem. The three managers understand that reducing the uncertainty for the overall system reliability estimate is of primary importance, but subsystem maintenance considerations are also important. Thus the goal is to find a best solution that simultaneously maximizes the estimation precision for the system and the mechanical and electrical subsystems, subject to the cost constraints and potentially different objective prioritizations.

To quantify the precision of the reliability estimates, the width of a 95% central credible interval for the reliability for each of these three nodes is used. The choice of the metrics on which to focus is a key element of successfully implementing multiple criteria optimization, since it ensures that the priorities of the study are appropriately quantified. Other options for measuring the uncertainty in reliability estimation such as the variance, standard deviation, or entropy could easily have been used to assess the precision, but the width of the credible interval was regarded as a familiar, easily interpreted option and matched how the results were typically presented.

The three managers proposed 60 allocations to consider, which are feasible given logistical constraints and represent good coverage of all the different potential allocations (see Table A in the supplementary materials II for the details of the allocations). The allocations all use the complete budget available, and the logistical constraints are dictated by data collection restrictions, such as the number of observations that can be obtained for each source and what fractions of the budget can be spent on each of the data sources. Often these constraints are imposed by system managers who only wish, or are able, to consider some of the total number of allocations possible. The allocations range from all full-system tests, through all subsystem level tests, to all component level tests, to mixtures of the different data sources. The goals of the optimization are first to determine the superior or admissible allocations, which balance the objectives for the system and subsystems from the 60 alternatives, and second, to compare these results to the global optimization Pareto front not subject to any logistical constraints in addition to the cost limit. Genetic algorithms are used to find the global front for all allocations within a fixed budget of 120 cost units. The decision making based on evaluating the tradeoffs and robustness for different prioritization of criteria is illustrated for the overall Pareto front, and compared with the most promising choices identified from the original 60 allocations.

The remainder of this article is structured as follows. In Section 2, we provide the background necessary for identifying a best allocation of new data when simultaneously optimizing multiple objectives. In Section 3, we consider Pareto front-based decision analysis based on the 60 logistically viable candidates identified by the three managers. We discuss using genetic algorithms and present a detailed Pareto front decision analysis to find a global best candidate allocation in Section 4. Section 5 concludes the paper with discussion of generalizations to the approach.

2. BACKGROUND

In this section, we present background on several building blocks required for selecting a best allocation of new data when simultaneously balancing multiple objectives using the Pareto front approach. We consider (1) resource allocation and (2) the Pareto front multiple objective optimization strategy.

2.1 Resource Allocation for System Reliability

Resource allocation is a form of sequential experimentation, where user-specified quantitative metrics are used to determine how to collect new data subject to a resource budget. In system assessment and management, the goal is often to more precisely estimate the system reliability. However, an additional goal may be more precise estimation of the reliability of other nodes. As with any sequential experiment selection, a decision must be made about how to collect the data without actually knowing what outcome will be observed. Hence, we wish to use our current knowledge of model parameters to inform us about what data results might be expected and how these results will influence the expected width of a credible interval for reliability. In this section we provide an overview of the approach proposed by Chapman, Morris, and Anderson-Cook (2012). This approach greatly streamlines the estimation of the expected credible interval width of the candidate allocations compared to the computationally intensive method in Anderson-Cook, Graves, and Hamada (2009).

Consider a vector of outcomes, $\boldsymbol{x}_{\text{new}} = (x_{\text{new},1}, x_{\text{new},2}, \ldots, x_{\text{new},k})$, from a candidate allocation $\boldsymbol{n}_{\text{new}} = (n_{\text{new},1}, n_{\text{new},2}, \ldots, n_{\text{new},k})$, where k is the total of the number of nodes, $\boldsymbol{n}_{\text{new}}$ is a vector with the proposed number of new observations for the nodes, and $\boldsymbol{x}_{\text{new}}$ represents a hypothetical number of successful tests on the nodes. Note that if no tests are proposed for a given node, then the entries corresponding to that node in $\boldsymbol{n}_{\text{new}}$ and $\boldsymbol{x}_{\text{new}}$ are 0.

Chapman, Morris, and Anderson-Cook (2012) considered the case where the interest is in improving the estimation precision of the system reliability, which they denoted as θ, where $\theta \in \boldsymbol{p}$, and noted that the updated posterior distribution of θ given the existing data and the specific hypothetical outcome $\boldsymbol{x}_{\text{new}}$ from the candidate allocation can be expressed as

$$f(\theta|\boldsymbol{x}, \boldsymbol{x}_{\text{new}}, \boldsymbol{n}, \boldsymbol{n}_{\text{new}}) = \frac{f(\boldsymbol{x}_{\text{new}}|\theta, \boldsymbol{n}, \boldsymbol{n}_{\text{new}})f(\theta|\boldsymbol{x}, \boldsymbol{n})}{f(\boldsymbol{x}_{\text{new}}|\boldsymbol{x}, \boldsymbol{n}, \boldsymbol{n}_{\text{new}})}, \quad (1)$$

where $\boldsymbol{x}$ is a vector summarizing the successes in the original data, $\boldsymbol{n}$ is a vector containing the number of tests performed on each node, and θ is the system reliability. While Chapman, Morris, and Anderson-Cook (2012) treated θ as the system reliability, it can represent the reliability of any individual node in the parameter vector $\boldsymbol{p}$.

Chapman, Morris, and Anderson-Cook (2012) noted that the three factors in (1) can be approximated using a large posterior sample generated from the MCMC-based Bayesian analysis of the existing data and combined to approximate (1) for a given outcome $\boldsymbol{x}_{\text{new}}$. To do this, first they noted that the posterior distribution of θ given the existing data, $f(\theta|\boldsymbol{x}, \boldsymbol{n})$, can be estimated

from the posterior draws of θ using kernel density estimation techniques over a grid of θ values.

To approximate $f(x_{\text{new}}|\theta, n, n_{\text{new}})$ and $f(x_{\text{new}}|x, n, n_{\text{new}})$ from (1), Chapman, Morris, and Anderson-Cook (2012) considered the pass/fail nature of the data. Under the assumption that tests on different nodes are performed independently, for each draw of p from the joint posterior distribution the probability of observing x_{new} from n_{new} is computed as the product, over all nodes, of the binomial probabilities of observing $x_{\text{new},i}$ successes out of $n_{\text{new},i}$ trials with probability $p_i^{(m)}$,

$$
f(X = x_{\text{new}}|n_{\text{new}}, p^{(m)})
= \prod_{i \in Q} \binom{n_{\text{new},i}}{x_{\text{new},i}} \left(p_i^{(m)}\right)^{x_{\text{new},i}} \left(1 - p_i^{(m)}\right)^{n_{\text{new},i} - x_{\text{new},i}},
$$

where $p^{(m)}$ represents the mth draw from the joint posterior distribution of reliabilities for the nodes. Averaging $f(x_{\text{new}}|n_{\text{new}}, p^{(m)})$ over the entire set of posterior θ draws provides an approximation of $f(x_{\text{new}}|x, n, n_{\text{new}})$, while averaging over subsets of these probabilities according to the grid of θ values allows for approximation of $f(x_{\text{new}}|\theta, n, n_{\text{new}})$ at a finite set of θ values. See Chapman, Morris, and Anderson-Cook (2012) for additional details on the approximation of the factors in (1).

The approximation of (1) is then used to estimate an uncertainty measure in the updated posterior distribution, such as the width of the 95% central credible interval. While Chapman, Morris, and Anderson-Cook (2012) focused only on system reliability, here we use this approach to estimate credible interval widths for the updated posterior distributions of the system and subsystems. Multiple hypothetical outcomes that are consistent with the existing data are randomly generated for each candidate allocation, and the credible interval width is estimated for each outcome; the average of these multiple estimated credible interval widths is reported as the expected credible interval width for the allocation. For a given candidate allocation, one such hypothetical outcome is generated by first randomly selecting one draw of node reliabilities from the joint posterior distribution. Suppose the mth posterior draw is randomly selected; then for the ith node being tested in the allocation ($i \in Q$), $x_{\text{new},i}$ is generated from a Binomial distribution with $n_{\text{new},i}$ trials and probability $p_i^{(m)}$. We use $t = 1000$ randomly generated hypothetical outcomes x_{new} for each candidate allocation. Finally, for each node of interest, the average of the t credible interval widths is reported as the expected credible interval width for the candidate allocation. We note that evaluating the same allocation multiple times can result in slightly different estimates because the randomly generated outcomes x_{new} will differ, but using $t = 1000$ for this example gives results consistent to the third significant digit.

This approach to quantifying the anticipated width of the credible interval for a new allocation is faster and requires smaller computational demands than previous approaches (such as Anderson-Cook, Graves, and Hamada 2009). Previous approaches required a new run of an MCMC algorithm for each hypothetical outcome, while this approach only requires the initial MCMC-based analysis. Thus, this approach is better suited for optimization problems in which a large number of candidate allocations are considered in the search process,

as presented in Section 4 where multiple genetic algorithm searches are conducted.

This approach to predicting the resulting changes in the characteristics of interest is based on the assumption that the model form used to describe the system is correct. For example, if the system structure was actually not a series system, then the original estimation could be biased, and future predictions of what new data would be expected could be misleading. In the resource allocation methodology, careful evaluation of the model assumptions is particularly important since new data are generated based on the specified form. The resulting improvements from this new data provide the ranking of different allocations, which could be misleading if the model form is incorrect. Hence, it is important to carefully evaluate the original model based on engineering understanding and model diagnostics to verify its appropriateness. See Anderson-Cook (2009a) for methods to assess series system assumptions.

2.2 Multiple Objective Pareto Front Optimization

We now consider Pareto front optimization in the context of design of experiments or resource allocation. In many decision-making scenarios, there are several simultaneous objectives to be considered and balanced. In the case of our reliability estimation scenario, the managers each want to minimize the width of the credible intervals of their nodes (system and two subsystems), but realize that no solution is likely to simultaneously achieve these three objectives. Hence, they need to reach a compromise satisfactory to each of them, after understanding what is possible.

Pareto optimization has been extensively used in many disciplines as a tool for optimizing multiple responses (Kasprzak and Lewis 2001; Gronwald, Hohm, and Hoffmann 2008; Trautmann and Mehnen 2009). However, only recently was the Pareto method introduced to the design of experiment paradigm (Lu, Anderson-Cook, and Robinson 2011, 2012). The desirability function (DF) approach of Derringer and Suich (1980), which has traditionally been used for design selection, converts the different criteria to a common desirability scale, typically between 0 (worst) and 1 (best) to create a single summary of the overall goodness of a design. Two common forms of the DF are additive (where the scaled criteria are combined as a weighted sum) and multiplicative (where the scaled criteria are combined into a product with the weights entering as exponents). Optimization using desirability functions does not directly consider the trade-offs between criteria, but rather a best design is identified with a "black box" algorithm that merges the different metrics into a single measure. As in our example, when different decision makers have different priorities, it is difficult to determine how these should be handled. The choice of which new data to collect can depend heavily on the user-specified relative importance of individual criteria, scaling schemes, and DF forms. Making decisions without understanding the potential impact from these important subjective choices can be risky.

Lu, Anderson-Cook, and Robinson (2011) adapted the Pareto front method to design of experiments and enhanced it with graphical tools to aid decision making. The method consists of two stages: (1) Pareto optimization, which is an objective step to assemble the set of contending designs where poor
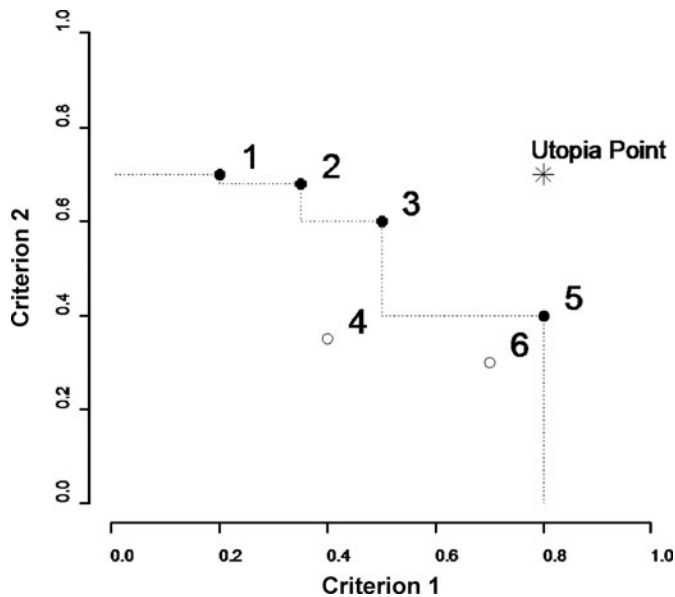
Figure 1. A simple illustration of a Pareto front for maximizing two criteria. The solution points are plotted on the original scale of the actual criterion values. The solid dots are points on the front. The open circles are points dominated by at least one point from the front. The asterisk symbol denotes the Utopia point.

designs (strictly inferior to others) are eliminated and no remaining choices are strictly better than others in the set for all criteria, and (2) the subjective Pareto decision analysis step that selects a subset of solutions from the Pareto front by examining tradeoffs between the criteria and robustness to different weights. The objective step of identifying the Pareto front allows the experimenter to see the full set of competing choices before considering the subjective aspects of the decision, and also allows for easy assessment of the robustness of results to these subjective choices.

One solution, here an allocation, is said to *Pareto dominate* another if it is at least as good for all criteria, and strictly better for at least one. This definition matches *admissibility* as defined for multiple criteria optimization of factorial blocking designs by Sun, Wu, and Chen (1997). The *Pareto set* is defined to contain solutions that are not Pareto dominated by others, and their corresponding criteria vectors form the Pareto front in the criterion space. Geometrically, the Pareto front is the set of points on the outer edge of the set of all solutions, closest to the ideal value for each objective. In the Pareto front literature, the *Utopia point* is defined as the vector that simultaneously has the best values observed for individual criteria in the solution space under consideration. Figure 1 illustrates a simple Pareto front based on maximizing two criteria. The points corresponding to individual potential solutions are plotted on the original scale based on the actual criteria values. Points 1, 2, 3, and 5 are identified to be on the Pareto front because no other point Pareto dominates them. Points 4 and 6 are not on the front because 4 is dominated by 3 and 5, and 6 is dominated by 5. The Utopia point is denoted by the asterisk symbol that corresponds to maximum values for both criteria. The dashed lines indicate where further points would be excluded from or included on the front. If more points are identified on the side of the dashed lines closer to the Utopia point, then they should be included for updating the Pareto front.

If any of the existing points on the current front are dominated by newly identified points, then they should be removed from the front. Any points identified on the side of the dashed lines farther away from the Utopia point should be excluded from the front. More discussion about Pareto-related concepts is available in a review article by Marler and Arora (2004).

If we choose from a fixed collection of candidates (such as the set of 60 logistically viable candidates), then the best value is determined based on the finite set of choices. If we are interested in a global solution, then the best value for each criterion should be found by the corresponding single-criterion optimization (which is equivalent to Pareto optimization based on an extreme weighting choice with all weight assigned for the single criterion; see Section 4.1 for more details). The Utopia point is useful as the "ideal" condition for calibrating individual design performance, but it is generally not attainable in practice. The Pareto set represents the collection from which a solution should be chosen, since any point in the interior of the cluster of solutions has at least one better alternative on the Pareto front.

The Utopia point approach (Lu, Anderson-Cook, and Robinson 2011) can reduce the Pareto set to a smaller manageable number of solutions by choosing designs closest in distance to the Utopia point, based on different weightings for combining the criteria. A common choice for measuring distances to the Utopia point is the L_1-norm, formulated as

$$\min_{\xi \in \Omega^*} \sum_{j=1}^{C} w_j |s_j(\xi) - s_j^0|, \qquad (2)$$

where Ω^* denotes the Pareto set of solutions, $s_j(\xi)$ is the jth objective function converted to a desirability scale between 0 and 1 for the jth criterion corresponding to solution ξ, and s_j^0 denotes the Utopia point value for the jth criterion on the same scale. Typically, we choose $s_j^0 = 1$. The w_j in (2) for $j \in \{1, 2, \cdots, C\}$ are the weights assigned to the C individual criteria. Each single-weight combination identifies one optimal solution from the Pareto front. However, not all solutions from the Pareto front will be optimal for at least one set of weights based on the chosen distance metric and scaling scheme. Hence, the Utopia point approach selects only a subset of the solutions from the Pareto front and further reduces the potential options to choose from for making a final decision. This Utopia point approach can identify the exact same set of solutions that will be selected by an additive DF, which is expressed as $DF(\xi, \boldsymbol{w}) = \sum_{j=1}^{C} w_j s_j(\xi)$, with the same weighting and scaling choices (Lu, Anderson-Cook, and Robinson 2011). From (2) we can see that the choice of a best solution depends on the weight, scaling, and form of the DF used. However, exploring different options for which allocation is best takes minimal computational effort once the Pareto front has been found. If the multiplicative DF, which penalizes poor performance on any single criterion more severely, is of interest, it can be converted to an additive form with a logarithmic transformation. In our case study, we consider simultaneously optimizing the improvements on precision of reliability estimates for the system, mechanical and electrical subsystems, hence we have $j \in \{S, M, E\}$. In addition, we choose to explore a fine grid of weight combinations spread over the entire weighting space

with all possible values being nonnegative multiples of 0.01 and summing to 1.

For solutions, which are best for at least one set of weights, each can be evaluated on three aspects: (1) performance for a particular set of weights that match the users' target for how the different objectives are valued, (2) the robustness of the choice of solution based on the range of weights close to those preferred by the user, and (3) the allocation's performance as measured by its desirability function score relative to the best possible score for a particular weighting choice. A final decision should be made by the user based on the priorities of the study after considering the tradeoffs and robustness of candidate solutions to different subjective choices. Compared to the standard DF method, the Pareto approach provides summaries of what range of solutions are available, more intuition about the relative performance of different solutions, and quantitative information for making a defensible coherent decision.

3. PARETO FRONT BASED ON INITIAL 60 ALLOCATIONS

Due to logistical constraints, 60 allocations were identified by the three managers as candidate choices. This set is quite diverse with varied fractions of the fixed budget for component, subsystem, and system level data. All of the allocations with their expected credible interval widths for the system and two subsystems reliability estimates are included in Table A in the supplementary materials II. A subset of the more promising of these allocations is shown in Table 2. Among the 60 allocations, the most precise estimate of either subsystem reliability is achieved by evenly splitting all new data between all components of the subsystem (Alloc. 41° for best improvement to the mechanical subsystem and 42° for best electrical improvement). The most precise estimate of the system reliability is achieved by spending equal amounts on all six components (Alloc. 52°). The estimate of the system reliability can be improved by up to 18.6% ($= 1 - 0.236/(0.722 - 0.432)$) relative to the original

Table 2. The 15 allocations on the Pareto front based on the original 60 allocations

Alloc #	n_{new} in the order of (M1, M2, E1, E2, E3, E4, M, E, S)	% reduction (M, E, S)
39°	(13,13,10,10,10,10,3,2,0)	(21.9,13.5,16.1)
41°	(30,30,0,0,0,0,0,0,0)	(**32.5**,0.4,9.6)
42°	(0,0,30,30,30,30,0,0,0)	(0.1,**26.0**,16.7)
43°	(15,15,15,15,15,15,0,0,0)	(21.5,16.3,17.8)
44°	(17,13,18,16,14,12,0,0,0)	(21.4,16.5,18.0)
45°	(20,40,0,0,0,0,0,0,0)	(30.7,0.4,9.2)
46°	(40,20,0,0,0,0,0,0,0)	(32.0,0.4,9.6)
48°	(0,0,40,40,20,20,0,0,0)	(0.3,25.3,16.2)
51°	(20,20,10,10,10,10,0,0,0)	(25.3,11.9,16.0)
52°	(10,10,20,20,20,20,0,0,0)	(15.3,20.1,**18.6**)
55°	(25,25,5,5,5,5,0,0,0)	(29.4,6.7,13.5)
56°	(30,20,5,5,5,5,0,0,0)	(29.4,6.8,13.5)
58°	(5,5,25,25,25,25,0,0,0)	(8.8,23.2,18.2)
59°	(10,15,30,20,10,10,0,0,0)	(17.8,18.3,18.0)
60°	(10,15,20,20,20,10,0,0,0)	(18.1,18.2,18.0)

NOTE: The maximum reduction for each criterion among original allocations is shown in bold.

data (using the width of the reliability credible interval for the system from Table 1). The estimates of reliability can improve by up to 32.5% and 26.0% for the mechanical and electrical subsystems, respectively.

Pairwise scatterplots of the percentage reduction in credible interval width for all 60 allocations are shown in Figure 2, with the 15 allocations on the overall three criteria Pareto front shown with solid circles (in italics in Table A). The ideal (unattainable) Utopia point allocation would simultaneously optimize all three criteria and would appear in the top right corner of each plot. First, it should be noted that some allocations perform very poorly and should not be considered further. The Pareto front allows us to remove $60 - 15 = 45$ allocations from further consideration, since these are dominated by at least one point on the front. The 15 allocations on the Pareto front represent the set of noninferior choices, with the relative improvement to the original data (% reduction) shown in Table 2. Depending on how the different criteria are valued, the best allocation from the original 60 allocations should be selected from these allocations (further discussed in Section 4.4).

Separate optimization searches based on each single criterion were conducted to find the best improvements possible given the cost restriction and used to calibrate the performance of the 60 allocations. Note that this is not required if one is only interested in selecting from the 60 allocations. However, it is recommended to help understand the actual performance of the given set of candidates relative to the best possible for individual criteria. Individually the best improvements possible are a $1 - 0.233/(0.722 - 0.432) = 19.4\%$ improvement for the system, and 33.0% and 27.0% for the mechanical and electrical subsystem reliabilities, respectively. Allocations 52°, 41°, and 42° have the best possible improvement for the three individual criteria among all the allocations in Table 2, with 18.6%, 32.5%, and 26% improvement for the system, mechanical and electrical subsystems, respectively. The proposed 60 allocations have a good coverage of the range of possible improvements and can produce near-optimal improvement on any criterion individually. This gives us partial information about performance, but we still do not know how this front compares to the best overall front on both the location of the front and its richness. Therefore, we expand the original problem to a more general scenario, where we seek the best allocations from the entire possible space subject only to the cost constraints. This provides a general methodology for finding the best overall solution.

4. GLOBAL PARETO FRONT

In this section, we consider how to construct a search algorithm to find the complete Pareto front for any allocation with a total cost within the 120-unit budget. After the overall front is found, we compare it to the front based on the original 60 allocations to help calibrate the performance of different candidate allocations. We also illustrate how to select a final allocation from the complete Pareto front based on the priorities of the study.

4.1 Genetic Algorithm for Populating the Pareto Front

To find the best possible Pareto front, an optimization search is needed to explore the allocation space for best performing
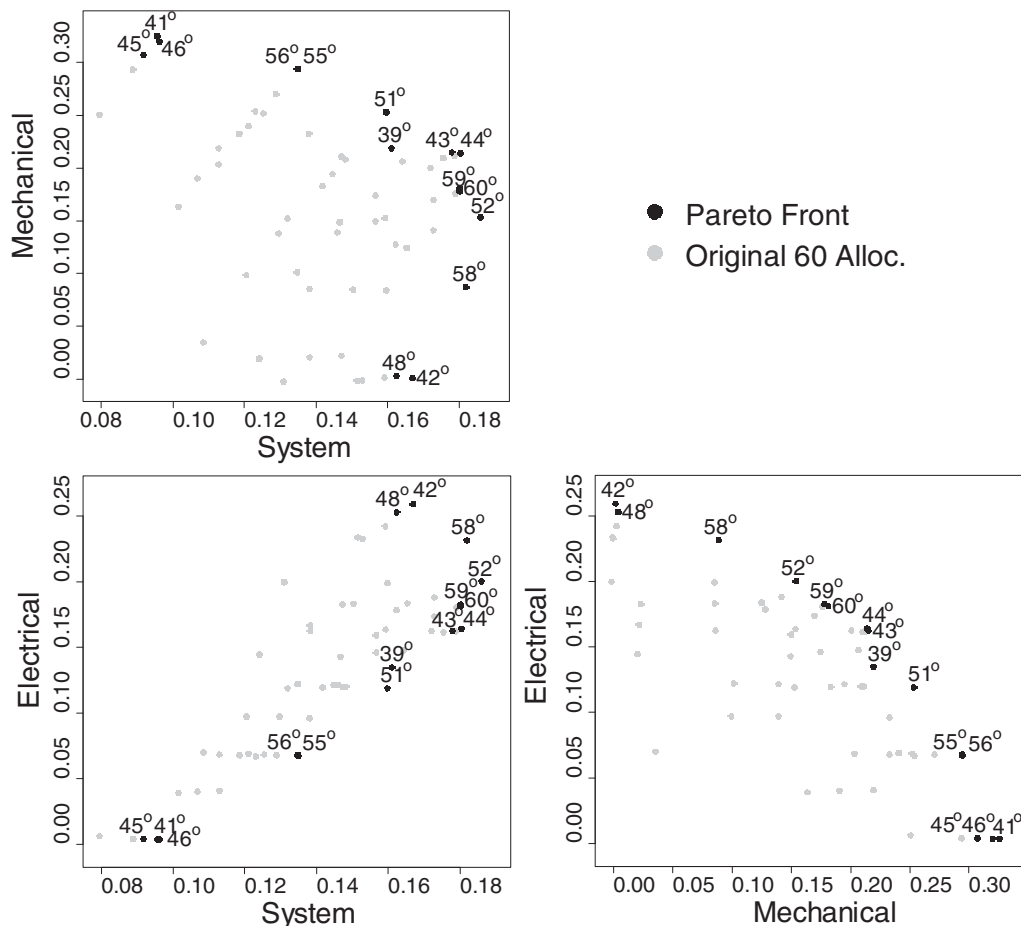
Figure 2. Pairwise scatterplots of the % reduction in estimated reliability credible interval width for the system, mechanical, and electrical subsystems for the original 60 allocations. The Pareto front is shown with black dots.

candidates. We use genetic algorithms (GAs) to generate and evaluate a large number of allocations as contenders for the front. A GA is a numerical optimization technique that mimics the evolutionary process to optimize a function of interest, often called a fitness measure (see De Jong 2006 and Hamada et al. 2008, Section 9.2). Iterations of a GA are known as generations, and at each generation, the candidate solutions comprising that generation compete for the ability to create new solutions. The basic steps of a GA are outlined next, followed by the details of the implementation of those steps in our application.

Step 1: Create and evaluate an initial population of N solutions. The solutions should then be ranked from best to worst fitness.
Step 2: Create N crossover/recombination solutions (multiple parent reproduction).
Step 3: Create N mutation solutions (single parent reproduction).
Step 4: Evaluate the $2N$ new solutions.
Step 5: Combine the $2N$ new solutions with the N parent solutions. Rank these solutions from best to worst fitness. Keep the top N of these solutions. This ensures that the best solution in Generation g will be no weaker than the strongest member in Generation $g-1$ (De Jong 2006).
Step 6: Repeat Steps 2–5 for a specified number of generations.

In our application, we combine the anticipated credible interval widths for the system and two subsystems using additive desirability to create the fitness measure to be optimized. We select the additive form of desirability because we do not wish to severely penalize allocations that may perform very poorly for one or more of the criteria; for these three objectives, additive desirability is expressed as $\mathrm{DF}(\xi, \boldsymbol{w}) = \sum_j w_j s_j(\xi)$, where, for $j \in \{M, E, S\}$, $s_j(\xi)$ is the credible interval width for the jth criterion corresponding to solution ξ and $\boldsymbol{w}$ is the weight vector with elements $w_j \geq 0$ subject to $\sum_j w_j = 1$.

Because the allocation that optimizes the combined fitness measure is highly dependent upon the choice of weights (w_j) considered, we use 21 runs of the genetic algorithm, each with a different weight vector $\boldsymbol{w}$, to optimize the additive desirability for multiple weights well spread throughout the complete weight combination space. The 21 sets of weights are chosen to consider weight values that are nonnegative multiples of 0.2 with the three entries summing to 1 for each set of weights. This 21 sets of weights form a small subset chosen from all possible weights (w_j) in (2) used in the Utopia point approach for selecting optimal designs across the entire weighting space. These different weights lead each of the GAs in different directions, thus allowing us to generate and consider a large number of diverse candidate allocations. A Pareto front is created based

on all of the candidates generated by each GA. Those 21 fronts are then pooled to create a final Pareto front.

To implement Step 1, we use the $N = 60$ candidate allocations identified by the system and subsystem managers as our initial population for all 21 weights considered. Alternatively, the initial population could be generated at random, although for good efficiency and convergence properties it is important to ensure good diversity across the allocation space. The three credible interval widths for these candidate allocations, as well as all candidate allocations considered by the GA, are estimated using the method outlined in Section 2.1. These criteria values are then used to calculate the combined fitness measure of the candidate allocation for the given weight vector.

In our implementation, rank-proportional selection is employed to choose parent allocations for crossover in Step 2. Specifically, the probability of selecting the rth-ranked candidate of the $N = 60$ solutions is $(N - r + 1)/[N(N + 1)/2]$ (Hamada et al. 2008, p. 322). Once two parents have been chosen, a child allocation is created. For example, suppose the two parents selected at random from the population are

Parent 1: (M1 = 8, M2 = 7, E1 = 0, E2 = 0, E3 = 0, E4 = 0, M = 5, E = 0, S = 5).
Parent 2: (M1 = 0, M2 = 0, E1 = 9, E2 = 8, E3 = 7, E4 = 6, M = 5, E = 12, S = 0).

Crossover proceeds by choosing at random (with equal probability) which parent will contribute at each location in the allocation. Suppose, the randomly generated parent contributions for the nine nodes are 2, 2, 1, 1, 2, 1, 2, 2, 1, and thus Parent 2 contributes to the first two nodes, etc. The resulting child allocation is then

Child: (M1 = 0, M2 = 0, E1 = 0, E2 = 0, E3 = 7, E4 = 0, M = 5, E = 12, S = 5).

The cost of this candidate allocation is 157 units, which exceeds the budget of 120 units. Hence, the candidate is repaired by randomly choosing one of the tested nodes and decreasing the number of tests allocated by one; this is repeated until the cost of the candidate allocation does not exceed the available budget. Similarly, if the cost of a child allocation is well below the budget, one of the tested nodes is selected at random and the number of tests allocated to that node is increased by 1; this is repeated until the cost of the allocation gets as close as possible to the budget while not exceeding it.

Next, in Step 3 each parent in the population is mutated once to create a set of $N = 60$ potential mutation allocations. For a single parent, mutation proceeds per the approach of Hamada et al. (2008, p. 323). Each entry of the parent's allocation is mutated with probability $\exp(-\mu g)$, where μ is the mutation rate and g is the generation number. Here, the mutation rate is specified as $\mu = 0.01$, though Hamada et al. (2008; p. 323) noted that changing this value has little effect on the performance of the algorithm. The mutation occurs such that when the ith entry of the allocation is mutated, the expected value of the mutation entry is approximately equal to the number of tests allocated to Node i by the parent allocation ($n_{\text{new},i}$) and the variability in the mutation entries decreases with each generation. For Node i, define G_i (and H_i) as the minimum (maximum) number of tests allowed on Node i. We use $G_i = 0$ and $H_i = \text{floor}(B/\phi_i)$,

where B is the budget amount and ϕ_i is the cost per observation for Node i, but these could be adapted to match logistical constraints.

Per Hamada et al. (2008; p. 323), for each $n_{\text{new},i}$, we define

$$z_i = \frac{n_{\text{new},i} - G_i}{H_i - G_i},$$

and

$$d_i = \log\left(\frac{z_i}{1 - z_i}\right) + (\text{Uniform}(0, 1) - 0.5) \times \sigma \times \exp(-\mu g),$$

where Uniform$(0, 1)$ represents a random draw from a uniform distribution and σ controls the rate at which the variation decreases with each generation and is specified here to be $\sigma = 1$ (Hamada et al. 2008; p. 323). The mutated sample size for Node i at Generation g is then computed as

$$n'_{\text{new},i} = \text{floor}\left(G_i + (H_i + 1 - G_i) \times \frac{\exp(d_i)}{1 + \exp(d_i)}\right).$$

This mutation is performed at each location in the parent allocation to form a new allocation.

In Step 4, the fitness measures of the 120 new allocations are computed. As a substep in our implementation, we take this opportunity to update the current Pareto front. Each candidate generated is compared to the current front. Existing allocations on the current front are replaced if any new candidates dominate them. If a candidate does not dominate any allocations on the front, but also is not dominated by anything on the current front, it is added to the front. If a new candidate is dominated by at least one allocation on the front, it is not added.

In Steps 5 and 6, the 120 new allocations are combined with the 60 allocations from the parent population, with only the top 60 allocations continuing to the next generation based on ranking the fitness measures. This cycle of reproduction and survival (Steps 2–6) was repeated for 60 generations.

Some additional comments on our implementation of the genetic algorithm: (1) In order to reduce the computation time and random variability due to the evaluation procedure, if a candidate is generated multiple times within the same GA it is only evaluated the first time it appears and the resulting credible interval widths for that allocation are used each subsequent time. (2) Occasionally, multiple runs of the GA would evaluate the same candidate allocation (e.g., runs that used similar weights). As noted in Section 2.1, evaluating the same candidate multiple times using the method of Chapman, Morris, and Anderson-Cook (2012) can result in slightly different estimates of the expected credible interval widths. Thus, the combined Pareto front uses the average value from multiple evaluations of a candidate allocation (when they exist) as the final approximation of the criteria values. The final fronts for the 21 weights were pooled and the overall Pareto front was created from all candidates considered.

The genetic algorithms for each of the 21 weights evaluated, on average, 3175 candidate allocations. The GAs for five of the weights evaluated fewer than 1000 candidate allocations, while the GAs for 10 of the weights required more than 4000 candidate allocations to be evaluated. The variability in the number of candidates evaluated can be attributed to how quickly the GAs converged; those that converged
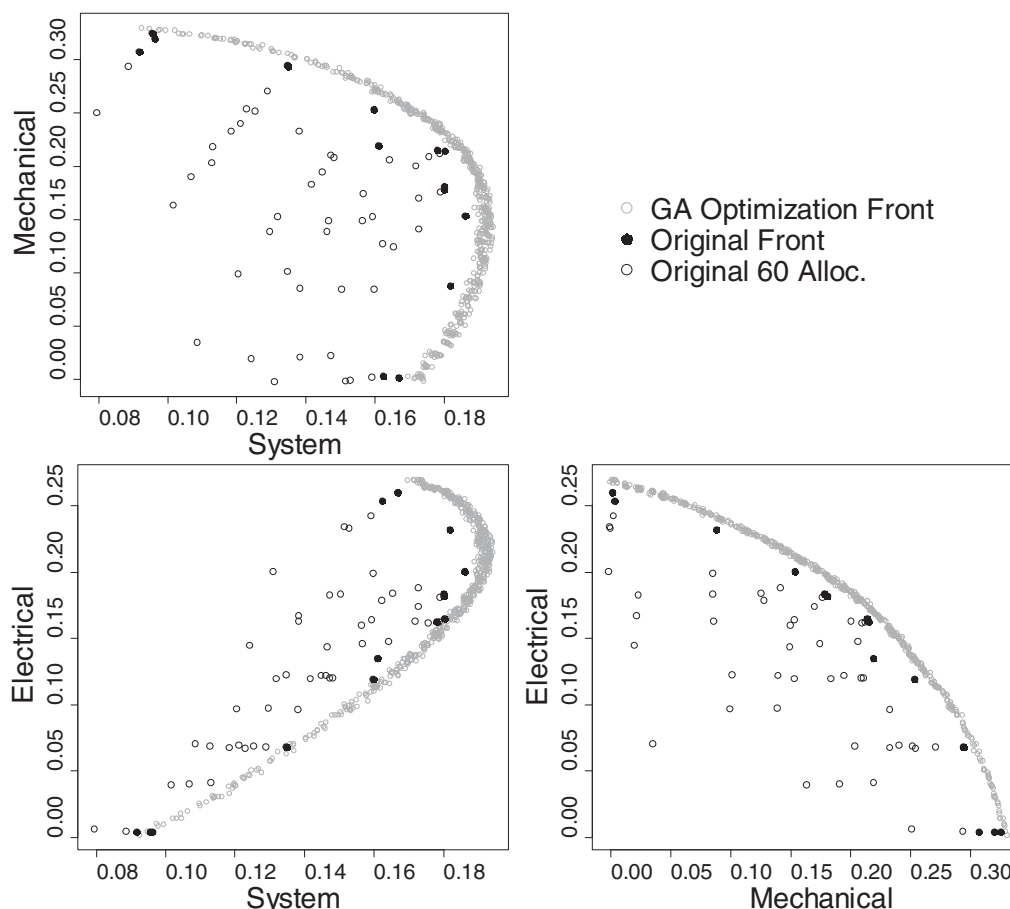
Figure 3. Pairwise scatterplots of the % reduction in the estimated reliability credible interval width for the system, mechanical and electrical subsystems for the original 60 allocations and its Pareto front and the Pareto front obtained by GA optimization.

slowly generated many more unique candidates to be evaluated, while those that converged quickly generated the same candidates many times. The genetic algorithm was implemented in R (R Development Core Team 2012) and interfaced with a program for implementing the method of Chapman, Morris, and Anderson-Cook (2012) written in C; this code is available in the supplementary materials. The genetic algorithms for the 21 weights were run in parallel on a high-performance computer, which is roughly equivalent to running each GA on its own machine with a 2.26 GHz processor and 6 GB of memory. The average run time for each of the 21 GAs was 58.8 hr, with run times ranging between 5 hr and 92 hr. While the time taken to create the overall Pareto front is substantial, it is still relatively short and inexpensive compared to the cost of collecting the actual data.

4.2 The Global Pareto Front

The global Pareto front found by the GA optimization search algorithm consists of 559 allocations, with the pairwise scatterplots of the % reduction in estimated reliability credible interval width for the system, mechanical, and electrical subsystems shown in Figure 3. Compared to the original front (black solid dots), the new front (dark gray circles) is consistently closer to the Utopia point and is more dense with more contending choices. None of the original 60 allocations are on the new front, which means the global optimization search has found strictly

better solutions than the choices available when constrained by logistics. Note that several allocations on the original front are relatively close to the final GA-based front and the ranges of both fronts are similar. Both of these factors suggest that the original proposed candidates gave good coverage of the possible choices, even though they are not strictly optimal.

The shape of the new front indicates a nearly proportional tradeoff between the two subsystems. This is intuitive since given the overall cost constraint, improving the estimation for one subsystem by collecting more data means less resources dedicated to the other. When we consider just two criteria, with the system and either of the two subsystems (the left panels in Figure 3), we see that the maximum improvement for the system occurs for moderate improvements of the subsystem. This is intuitive since to improve the system estimate we need to obtain data that also improve the subsystem estimate. In fact, as we traverse these pairwise Pareto front between potential solutions, improving the subsystem estimates initially is associated with nearly linear improvement in system reliability estimation. However, after reaching the optimum, a little sacrifice in system gain allows disproportionate gains in the subsystems. The mechanical subsystem shows more tradeoffs with the system than the electrical subsystem, which is evidenced by the fact that the front for the system and the electrical subsystem is closer to the Utopia point than the front for the system and the mechanical subsystem.
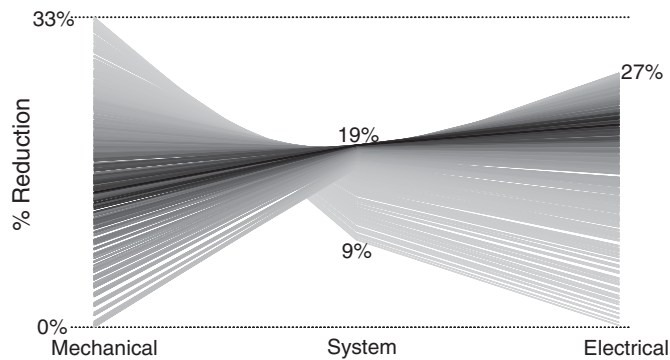
Figure 4. Parallel plot for the Pareto front obtained by GA optimization. The darker lines represent allocations with higher improvement on estimation of system reliability.

Figure 4 shows a parallel plot for the percentage reduction in the width of the uncertainty intervals possible for the three criteria based on all of the solutions on the global Pareto front found by using the GA. All criteria are plotted on the same scale (with the % reduction being mapped from 0% to 33%), with each line representing a single allocation. Because system reliability is most important, we use darker lines to represent closer to improvement on its estimation. The amount of crossing of lines between adjacent criteria indicates the amount of tradeoffs between them. We can see substantially more tradeoffs between the mechanical subsystem and the system than for the electrical subsystem and the system. The parallel plot scales well to higher dimensions, and the use of gray-scale for the lines based on different prioritization of criteria is effective to reveal the interrelationship between criteria.

4.3   Decision Based on the Global Pareto Front

The identification of the Pareto front provides an objective set of contending allocations. However, ultimately the managers must decide on a single allocation to collect the new data. Next, we illustrate a strategy for how to reduce the number of choices by intentionally using the priorities of the study to subjectively influence the final decision through an informed and justifiable approach.

The first step for reducing the number of choices is to use the Utopia point approach to identify allocations that are best for at least one weighting combination of the criteria based on a chosen metric form and scaling scheme for criteria. In our application, since all three criteria are directly comparable by using the % reduction scale, we choose 0% as the lower bound (mapped to 0) and maximum reduction from any of the three criteria, (i.e., 33%) is mapped to 1 on the desirability scale, and use the same scaling for all three criteria. The mechanical subsystem can achieve the 33% maximum improvement among all three criteria and have observed values expanding over the full range of the plot scale. The system and electrical subsystem have the observed ranges of (9%, 19%)/33% and (0, 27%)/33%, respectively. Also, we use the L_1-norm in (2) for the distance metric (equivalent to the additive DF) since we do not want to unduly penalize allocations that perform poorly on one criterion. The results of scaling and the form of the distance metric
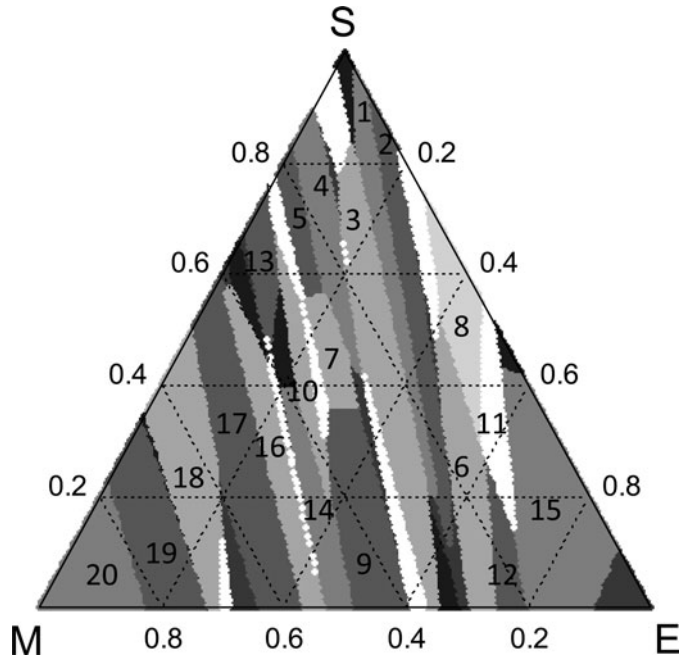

Figure 5. Mixture plot for allocations selected based on the L_1-norm for different weightings of the criteria. Allocations corresponding to at least 1.5% of the total simplex area are numbered with smaller number for more improvement on precision of system reliability estimation.

do influence subsequent results, so if the data collector is uncertain about these choices, a sensitivity analysis is recommended. This requires little extra effort once the Pareto front has been identified.

Figure 5 shows the mixture plot of the best allocations for different weightings of the three criteria (Lu, Anderson-Cook, and Robinson 2011). Every point in the triangle corresponds to a weight combination with the sum of the three weights equaling one. Hence, the vertices and the edges correspond to optimizing based on a single criterion and two of the three criteria, respectively. The weight combinations corresponding to the same optimal design are displayed in the same region in the same grayscale. There are 51 allocations selected by the Utopia point approach from the 559 on the Pareto front, and 20 allocations with better robustness (corresponding to at least 1.5% of the total simplex area) are numbered based on % reduction in system credible interval width (a smaller number indicates more improvement). We can see that the mixture plot contains many small regions (shown in different gray-scale) without any dominating choice corresponding to a large region. This is not surprising since the front is rich and adjacent solutions are quite closely positioned and perform similarly. However, this streamlined process effectively selects a smaller manageable set of more promising allocations (3.6% = 20/559 of the allocations on the front) to be further compared using individual performance and robustness to weight uncertainty.

The 20 numbered allocations can now be examined more closely, since there are a manageable number of choices to consider. Allocation 1 has the second best system improvement. However, compared to the global optimum based on only system reliability (the allocation at the top corner corresponding to putting 100% weight for system reliability), it is a more

robust choice (bigger area) with only 0.025% less improvement than the best allocation for system improvement. The weights for which this allocation is best correspond to the system being weighted most and improvements to the electrical subsystem estimation are considered slightly more important than the mechanical subsystem. Allocation 3 corresponds to the third biggest area with relatively high emphasis on system improvement (system weighted no more than 80%) and the electrical subsystem considered slightly more important than the mechanical subsystem. Allocation 4 is best when the system is weighted no less than 40% and the mechanical subsystem is valued slightly more than the electrical subsystem, which also is best when the system is weighted 50% and the remaining weight is split evenly between the two subsystems. Allocation 7 is the most promising choice when system is weighted between 35% and 55% and the mechanical subsystem is weighted slightly more than the electrical subsystem. Allocation 15 (with the biggest area in the simplex) should be selected when the electrical subsystem is considered much more important than the mechanical subsystem, while Allocation 18 (or 17) is best when the mechanical subsystem is weighted more than the electrical subsystem. Note that there are many small regions shown as strips running between the system and electrical subsystem corners of the simplex when mechanical subsystem is weighted less, but as the weight increases the regions become larger. This pattern is related to the greater tradeoff identified between the mechanical subsystem and other two criteria in Figure 4.

Figure 6 shows the trade-off plot (Lu, Anderson-Cook, and Robinson 2011) for the 51 allocations shown in Figure 5. This plot shows the criteria values (% reduction) for optimal allocations selected by the Utopia point approach (all allocations that are optimal for at least one weight combination shown in Figure 5). The three sets of symbols show the criterion values for a given allocation, with the 20 numbered allocations shown in larger black symbols and the remaining unnumbered ones in smaller gray symbols. The allocations are sorted from left to right based on best to worst system improvement. Note that the range of system improvement is substantially narrower than the two subsystems. Second, there is less tradeoff between the
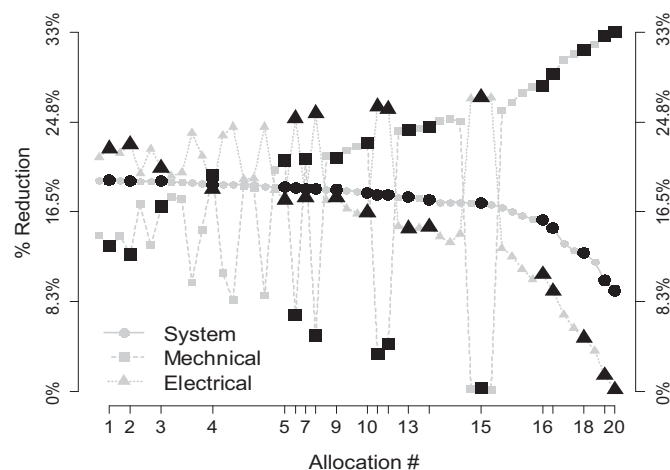

Figure 6. Tradeoff plot for allocations selected based on the L_1-norm for different weightings of the criteria. The numbered allocations in Figure 5 are highlighted in black symbols.

system and the electrical subsystem, especially for allocations with lower system estimation improvement (close to the right). Allocations 3, 4, and 7, all represent choices with good balanced performance for all three criteria with an emphasis on good system improvement in precision. Allocations 1, 15, and 18 each represent a choice based on strongly emphasizing one of the three criteria. Allocations close to the discussed ones are expected to have similar performance. A smaller set of most promising allocations should be selected based on Figures 5 and 6 and compared over the region of interest for weight combinations that match the particular goals of data collection.

The final step before selecting a single allocation is to compare the performance of a few promising allocations relative to the best possible for different weights with the synthesized efficiency plots (Lu and Anderson-Cook 2012) in Figure 7. Synthesized efficiency provides a measure of relative performance compared to the global optimal performance for a given weighting. When using desirability functions to combine the three criteria into a single metric, the synthesized efficiency of Solution ξ at the weight vector $\mathbf{w}$, relative to the global optimal solution for weight $\mathbf{w}$, is defined as

$$\mathrm{SE}(\xi, \mathbf{w}) = \frac{\mathrm{DF}(\xi, \mathbf{w})}{\max_{\xi}(\mathrm{DF}(\xi, \mathbf{w}))}.$$

The white-gray-black coloring scale is used to represent high to low synthesized efficiency. Allocations 3, 4, and 7 each have white or light gray color for the majority of the weighting space (with about 83%, 85%, and 86% of the weighting area with synthesized efficiency above 80%, respectively). The lowest efficiency is 52%, 60%, and 65% for Allocations 3, 4, and 7, respectively. This shows that these allocations perform well across a broad range of different prioritizations of the different criteria. Allocation 1 has a larger white region near the system corner than the above three, but across the entire simplex has a smaller region (77%) with at least 80% efficiency as well as poorer lowest efficiency (41%). Allocation 15 has a white region near the electrical corner but as low as 0% efficiency at the mechanical corner. Similarly, Allocation 18 is at least 95% efficient for a relatively large area close to the mechanical corner but has as low as 18% efficiency at the electrical corner. If the data collector has a focused weight region of interest, a best allocation can be selected based on its performance in that region.

To summarize across the weighting region of interest, we develop a new graphical summary, the fraction of weight space (FWS) plot that was adapted from the fraction of design space (FDS) plot (Zahran, Anderson-Cook, and Myers 2003). It provides an overall quantitative summary of individual design performance across the weighting space and, hence, offers easy comparison between competing choices. As a complement to the synthesized efficiency plot, which shows where in the weighting space an individual solution performs relative to the best possible, the FWS plot summarizes the solution performance for the entire weighting space regardless of its dimension with a single curve and allows for easy assessment of overall performance (best, worst, and robustness/stability) of individual solutions and easy comparison among many competing choices. Better solutions have generally large synthesized efficiency values and flat curves across the entire region of the fraction of weighting
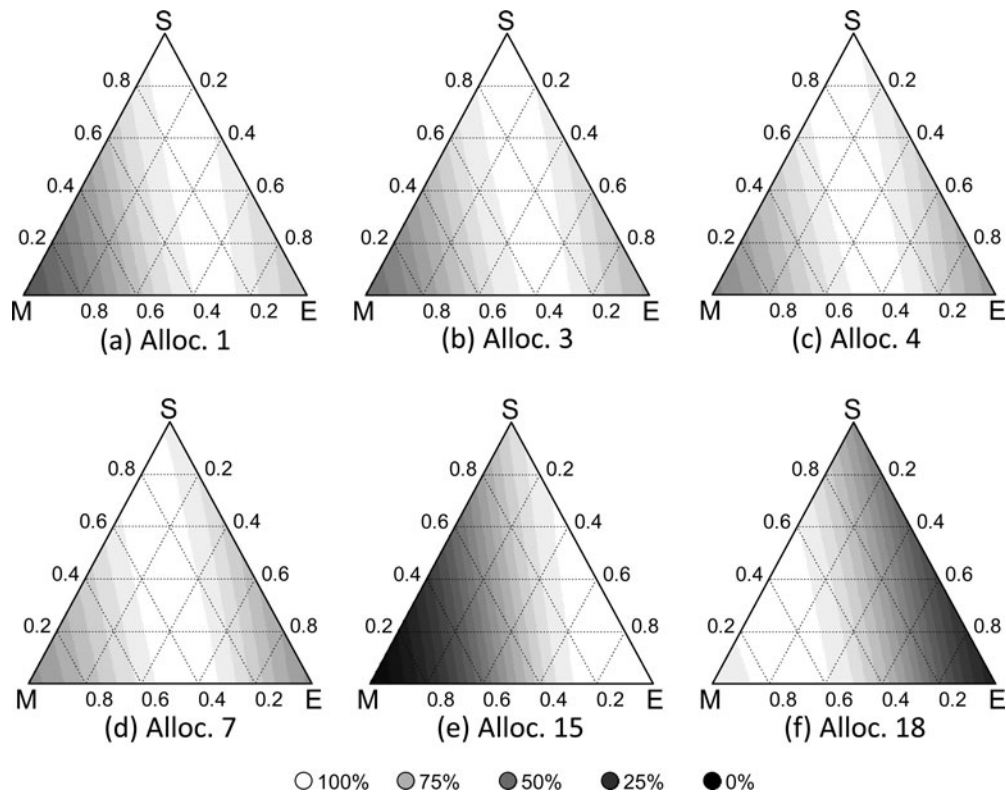
Figure 7. Synthesized efficiency plots for some of the allocations from Figure 5.

space. Here we consider the entire possible weighting space without specifying a more focused region of interest; however, we could easily adapt the summary to a portion of the weight space to match the decision makers' priorities. Figure 8 shows the FWS plot for the allocations in Figure 7. The line for each design shows the fraction of the weighting space for which the synthesized efficiency is at least as large as a certain percentage. A discrete approximation of this summary can be built from the information in Figure 7. For each design, we have a vector of synthesized efficiency values relative to the best possible for each weight combination with fine coverage of the entire weighting space. Then we sort the efficiency values in descending order and extract a list of distinct values in the same order. For each distinct value, we calculate the fraction of entries (weight combinations) in the sorted efficiency vector at least as large as that value.

From Figure 8, all six allocations have close to 100% synthesized efficiency for around 17% of the weighting space. Allocations 3, 4, and 7 have overlaid lines for about 65% of the weighting space, and then for Fraction of Weighting Space values greater than 65%, Allocation 7 has consistently slightly better values than Allocation 4, and both are better than Allocation 3. Allocation 1 has slightly lower synthesized efficiency than Allocations 3, 4, and 7 for up to 70% of weighting space, and the difference becomes more pronounced for the remaining weighting space. Allocations 15 and 18 have substantially lower synthesized efficiency than the others for the majority of the weighting space, with Allocation 18 being worst among the six with lowest efficiency around 0%. The FWS plot provides an effective approach for quantifying the per-

formance of individual choices over the weighting region, is very for any number of criteria, and, hence, provides a straightforward and scalable comparison between candidate decision choices.

Table 3 shows the 20 numbered allocations from Figure 5 with the number of new observations to collect for each data source. All 20 allocations have no observations for any of the system and subsystems. The same pattern was also seen from the allocations on the Pareto front based on the original 60. It is a result of the cost structure for the system. Since having one observation for
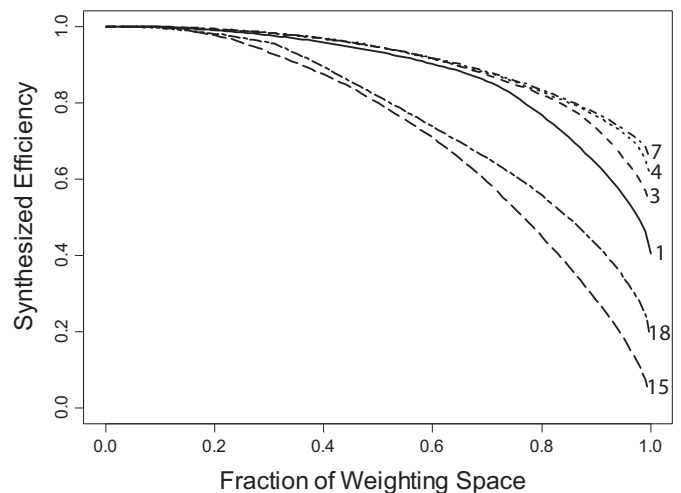

Figure 8. Fraction of weight space (FWS) plot for some of the allocations from Figure 7.

Table 3. The size of different data sources and criteria values for the numbered allocations in Figure 5

Alloc #	n_{new} in the order of (M1, M2, E1, E2, E3, E4, M, E, S)	% reduction (M, E, S)
1	(10,6,32,15,21,20,0,0,0)	(13.4,22.3,19.4)
2	(9,6,32,15,23,20,0,0,0)	(12.5,22.7,19.4)
3	(12,9,29,12,18,19,0,0,0)	(17.0,20.6,19.3)
4	(14,12,29,10,16,13,0,0,0)	(19.9,18.6,19.0)
5	(16,13,23,11,13,15,0,0,0)	(21.2,17.6,18.8)
6	(6,1,37,15,32,22,0,0,0)	(7.0,25.1,18.7)
7	(15,14,23,10,17,12,0,0,0)	(21.3,17.8,18.6)
8	(5,0,38,18,26,28,0,0,0)	(5.1,25.5,18.6)
9	(17,12,25,9,17,11,0,0,0)	(21.4,17.8,18.6)
10	(18,14,21,9,14,12,0,0,0)	(22.8,16.4,18.3)
11	(3,0,39,17,30,28,0,0,0)	(3.4,26.2,18.1)
12	(4,0,39,19,29,25,0,0,0)	(4.4,26.0,18.1)
13	(19,16,25,7,9,9,0,0,0)	(24.0,15.0,17.9)
14	(18,17,24,7,10,9,0,0,0)	(24.3,15.1,17.6)
15	(0,0,45,19,29,27,0,0,0)	(0.3,27.0,17.3)
16	(22,22,13,5,8,6,0,0,0)	(28.1,10.7,15.8)
17	(25,22,17,0,6,3,0,0,0)	(29.2,9.2,15.0)
18	(29,25,10,0,0,2,0,0,0)	(31.4,4.9,12.8)
19	(30,29,2,0,0,0,0,0,0)	(32.7,1.5,10.2)
20	(32,28,0,0,0,0,0,0,0)	(33.0,0.1,9.2)

Figure 9. Fraction of weight space (FWS) plot for comparison between best allocations from the original 60 and the GA optimization.

all components is cheaper than for both subsystems or for the system, it is intuitive to have all component observations in the new data to maximally improve the reliability estimates.

Among the six allocations in Figures 7 and 8, Allocations 1 and 3 use most of the resources (about 75% of the total cost) to collect electrical component data and have the most gains in improving system and electrical subsystem reliability estimates. Allocations 4 and 7 distribute the cost more evenly between the two subsystems' components to achieve good performance for all reliability estimates of interest. Allocation 15 uses all the resources for collecting electrical component data with the most improvement in the electrical subsystem estimation and the least in the mechanical subsystem estimation. Allocation 18 selects the majority of the new data (about 90% of total cost) for improving the mechanical subsystem reliability estimate and, hence, less gain in system and electrical subsystem estimation. The observed pattern of the most promising allocations matches with our understanding of the tradeoffs between the three criteria obtained from Figures 3 and 4.

The choice of a final "optimal" allocation should involve understanding the available choices, observing tradeoffs between the criteria of interest for promising solutions, and a subjective determination of a preferred weight associated with each criterion. The goal of the Pareto front approach is to provide the set of sensible alternatives in the objective phase, and then to provide numerical and graphical summaries to equip the decision maker to decide what is best for the study goals.

4.4 Comparison Between Constrained Choices and Global Optimization

Next we compare the best choices from the GA global optimization search with the best solutions from the original 60
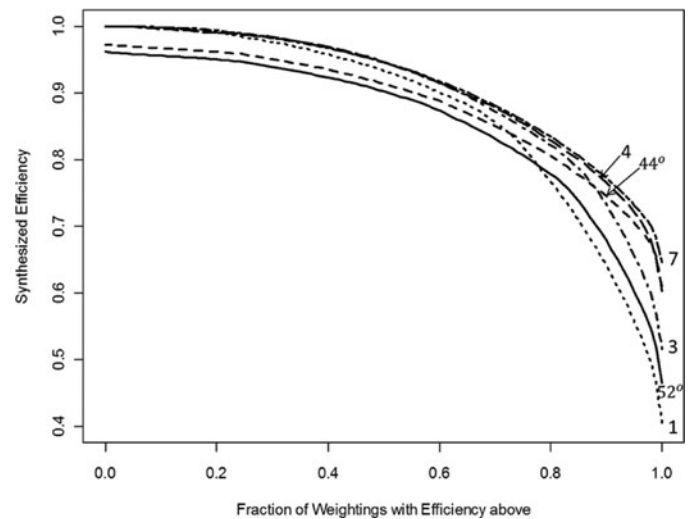
allocations subject to the logistic constraints. Having found the Pareto front based on the original 60 allocations in Figure 2, similar decision making can be used to select allocations best suited to the priorities of the study. The graphical summaries for examining robustness and tradeoffs between promising choices for different weightings of criteria are provided in supplementary materials III. Among the 15 allocations on the front, 6 are best for substantial proportions of the weight space. Allocation 52° would be preferred when system precision is most important, as it reduces the width of the system reliability credible interval the most. Allocation 44° sacrifices 0.6% and 3.6% on the system and electrical subsystem estimation, respectively, relative to 52° but achieves a 6.1% gain in the mechanical subsystem, which is best when the system is valued most and the mechanical is considered slightly more important than the electrical subsystem, as well as when the system is down weighted to be considered as important as the two subsystems. It is the most robust choice (best for biggest weighting area) and has consistently high synthesized efficiency when summarized over the majority of the entire weighting space. Allocation 58° has good improvement for the system, and might be considered when improving the electrical subsystem estimation is more heavily weighted. The remaining allocations (56°, 42°, 41°) do not perform well for improving system reliability precision, and, hence, are not likely considered as promising candidates.

Figure 9 shows how the top two original allocations (44° and 52°) compare to the most promising choices identified from the GA optimization (Allocations 1, 3, 4, and 7). Note the best possible of the original candidates are now calibrated based on what is possible from the overall GA front. We can see that Allocation 7 from the GA optimization has consistently better performance (at least 5% more synthesized efficiency for majority of the weighting space) than the two best original allocations. Allocation 52° tends to have substantially worse performance than all GA solutions in Figure 9 except for Allocation 1 when considering the worst efficiencies of the 20% weighting space. Allocation 44° is also consistently outperformed by GA solutions for the majority of the weighting space (around 72% of the

higher efficiency region). It is better than Allocations 1, 3, and 4 only when around 28%, 13%, and 2% of the lowest efficiency area is of interest, respectively. Therefore, the GA optimization finds allocations with better overall performance for most of the possible weightings of three criteria. Based on this additional information, the decision maker is now in a better position to assess the value of the originally proposed allocations, and also to consider whether it might be possible to relax the logistical constraints to consider any of the best allocations from the overall GA front.

5.   CONCLUSIONS AND DISCUSSIONS

This article adapts the Pareto front approach to simultaneously optimize multiple objectives in the resource allocation context. In current applications, many decisions about how resources should be spent are not based on formally evaluating all possible choices. This represents a lost opportunity for getting the most information out of precious resources. We recommend a streamlined decision-making process to (1) formally identify the goals of the data collection, (2) choose appropriate metrics for quantifying the objectives to match the study goals, (3) understand the scope of choices under consideration (based on logistical constraints or the global Pareto front), (4) identify the Pareto optimal choices from either a finite set of candidates (predetermined by the logistical constraints) or through a global optimization search algorithm, and (5) use numerical and graphical summaries to formally look at the tradeoffs between the options from the Pareto front to understand their relative merits and make a justifiable decision. The method can flexibly handle different numbers of criteria with different quantitative measures and different levels of richness of the scope of choices. If this process is followed, higher quality decisions can be made to achieve maximum benefit from limited resources, and different parties in decision making are more likely to achieve consensus about the merits of the choices made.

In studying the multiple response optimization literature and reflecting with our own practical experience, we believe that all of us do not make complex decisions with multiple objectives in the same way. However, we do believe that the Pareto front approach is rational, defensible, and flexible to adapt to different decision-making scenarios. Regardless of the details of the approach, it is universally helpful to objectively eliminate "noncontenders." This makes the construction of the Pareto front a logical first step. The parallel plot of the complete front is then effective for showing the tradeoffs and interrelationship between criteria.

Once we have identified the set of "contenders," it is helpful to identify a smaller subset to examine more closely with the mixture and tradeoff plots. These highlight which choices perform well depending on how we value the different objectives. The graphical tools provide an easy way to visualize the nature of the relationship between choices, quantify the tradeoffs between the alternatives, and facilitate discussion between decision makers who may have different priorities. Finally, for this smaller number of choices, it is helpful to look at how much we are sacrificing over the best available for a given set of weightings of the criteria with the synthesized efficiency plot. The new FWS plot allows a comparison of multiple choices by summarizing

individual performance across the weighting space. It can be used for any number of criteria and is adaptable for nonuniform weighting preference if specified by the decision maker.

This article showcases two typical scenarios that practitioners might encounter: one is the case when a set of candidate choices is limited by logistical constraints and the final decision has to be made from the given candidate set; the other scenario allows all possible choices to be considered. In the decision-making process, the two cases differ in the step of finding the Pareto front. The unconstrained front has to be identified by a global optimization search algorithm. In our application, a generic algorithm (GA), which conducts multiple parallel searches using a collection of prespecified weight combinations, is developed, and the Pareto front is populated from the individual searches and then combined to produce the overall front. The initial population for the GA can be either user-identified candidates or a set of choices with good coverage of the entire solution space.

Note that the best solutions identified in this article rely on the assumed cost structure. With system and subsystem data being more expensive, the best allocations were comprised of only component-level data. This pattern can change substantially if different cost structures exist. For example, if the system and subsystem data are cheaper to obtain, then the best allocations will likely be comprised of system and/or subsystem data to balance between cost-effectiveness and avoid the ambiguity of information from an observed system failure. Even though our case study explored only a scenario with a fixed overall cost, the adaptation to variable cost is straightforward by changing the cost constraint in the GA search algorithm to the allowable range or repeating the optimization for different values of fixed cost and combining the obtained fronts. The best allocation also depends on the starting point of the resource allocation problem, as the amount of each source of data already obtained and the reliability of the different system parts both play important roles in determining what new data are most valuable.

This case study illustrates the application of the Pareto approach for improved decision making in the resource allocation context using a simple single-use series system with nonaging parts and observed pass/fail data. However, the Pareto front approach for resource allocation problems with multiple criteria could be adapted to more complicated systems with different structures and more data types available, such as degradation data from multiple levels for aging systems.

SUPPLEMENTARY MATERIALS

Supplementary materials for this article are available online, and include four parts:

I. **Implementation of the Metropolis-Hastings Algorithm** provides a brief outline of our implementation of the Metropolis-Hastings algorithm (Chib and Greenberg 1995) used to generate a sample from the joint posterior distribution of node reliabilities.

II. **The original 60 allocations** has a table of the expected widths of credible intervals for the estimates of system and subsystem reliability for the original 60 allocations identified subject to the logistical constraints with a total cost of 120 units.

III. **Decision based on the original 60 allocations** shows the graphical summaries obtained based on the original 60 allocations subject to the logistical and cost constraints.

IV. **Computer code:** A separate zip file with the computer code used for implementing the method of Chapman, Morris, and Anderson-Cook (2012), the GA search, and finding the global Pareto front based on separate GA searches is provided. The code is available in two versions: one is the version that was actually run in Linux to generate the results shown in the main article with some portion written in C and the remaining written in R; the other is a version written entirely in R and is usable on any platform.

ACKNOWLEDGMENTS

[Received May 2012. Revised July 2013.]

REFERENCES

Anderson-Cook, C. M. (2009a), "Evaluating the Series or Parallel Structure Assumption for System Reliability," *Quality Engineering*, 21, 88–95. [476]
——— (2009b), "Opportunities and Issues in Multiple Data Type Meta-Analyses," *Quality Engineering*, 21, 241–261. [473]
Anderson-Cook, C. M., Graves, T. L., Hamada, M., Hengartner, N., Johnson, V. E., Reese, C. S., and Wilson, A. G. (2007), "Bayesian Stockpile Reliability Methodology for Complex Systems," *Journal of the Military Operations Research Society*, 12, 25–37. [473]
Anderson-Cook, C. M., Graves, T. L., and Hamada, M. S. (2009), "Resource Allocation for Reliability of Complex Systems With Aging Components," *Quality and Reliability Engineering International*, 5, 481–494. [474,475,476]
Chapman, J. L., Morris, M. D., and Anderson-Cook, C. M. (2012), "Computationally Efficient Comparison of Experimental Designs for System Reliability Studies With Binomial Data," *Technometrics*, 54, 410–424. [474,475,476,480,487]
Chib, S., and Greenberg, E. (1995), "Understanding the Metropolis-Hastings Algorithm," *The American Statistician*, 49, 327–335. [486]
De Jong, K. A. (2006), *Evolutionary Computing: A Unified Approach*, Cambridge, MA: MIT Press. [479]
Derringer, G., and Suich, R. (1980), "Simultaneous Optimization of Several Response Variable," *Journal of Quality Technology*, 12, 214–219. [476]
Gronwald, W., Hohm, T., and Hoffmann, D. (2008), "Evolutionary Pareto-Optimization of Stably Folding Peptides," *BMC-Bioinformatics*, 9, 109. [476]
Hamada, M., Wilson, A. G., Reese, C. S., and Martz, H. F. (2008), *Bayesian Reliability*, New York: Springer. [473,479,480]
Johnson, V. E., Graves, T. L., Hamada, M., and Reese, C. S. (2003), "A Hierarchical Model for Estimating the Reliability of Complex Systems" (with discussion), in *Bayesian Statistics 7*, Oxford University Press. [473]
Kasprzak, E. M., and Lewis, K. E. (2001), "Pareto Analysis in Multiobjective Optimization Using the Collinearity Theorem and Scaling Method," *Structural Multidisciplinary Optimization*, 22, 208–218. [476]
Lu, L., and Anderson-Cook, C. M. (2012), "Rethinking the Optimal Response Surface Design for a First-Order Model With Two-Factor Interactions, When Protecting Against Curvature," *Quality Engineering*, 24, 404–422. [483]
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, 353–365. [476,477,482,483]
——— (2012), "A Case Study to Demonstrate Pareto Frontiers for Selecting a Best Response Surface Design With Simultaneously Optimizing Multiple Criteria," *Applied Stochastic Models in Business and Industry*, 28, 206–221. [476]
Marler, R. T., and Arora, J. S. (2004), "Survey of Multi-Objective Optimization Methods for Engineering," *Structural Multidisciplinary Optimization*, 26, 369–395. [477]
Martz, H. F., and Waller, R. A. (1990), "Bayesian Reliability Analysis of Series Systems of Complex Series/Parallel Systems of Binomial Subsystems and Components," *Technometrics*, 32, 407–416. [473]
Martz, H. F., Waller, R. A., and Fickas, E. T. (1988), "Bayesian Reliability Analysis of Series Systems of Binomial Subsystems and Components," *Technometrics*, 30, 143–154. [473]
R Development Core Team (2012), *R: A Language and Environment for Statistical Computing,* Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0. Available at *http://www.R-project.org/*. [481]
Sun, D. X., Wu, C. F. J., and Chen, Y. (1997), "Optimal Blocking Schemes for $2n$ and $2n–p$ Designs," *Technometrics*, 39, 298–307. [477]
Trautmann, H., and Mehnen, J. (2009), "Preference-Based Pareto Optimization in Certain and Noisy Environments," *Engineering Optimization*, 41, 23–38. [476]
Zahran, A., Anderson-Cook, C. M., and Myers, R. H. (2003), "Fraction of Design Space to Assess Prediction Capability of Response Surface Designs," *Journal of Quality Technology*, 35, 377–386. [483]