Optimizing in a Complex World
Christine M. Anderson-Cook, Statistical Sciences Group, Los Alamos National Laboratory

Abstract
As applied statisticians increasingly participate as active members of problem-solving and decision-making teams, our role continues to evolve. Historically, we may have been seen as those who can help with data collection strategies or answer a specific question from a set of data. Nowadays, we are or strive to be more deeply evolved throughout the entire problem-solving process. An emerging role is to provide a set of leading choices from which subject matter experts and managers can choose to make informed decisions. A key to success is to provide vehicles for understanding the trade-offs between candidates and interpreting the merits of each choice in the context of the decision-makers' trade-offs. To achieve this objective, it is helpful to be able (a) to help subject matter experts identify quantitative criteria which match their priorities, (b) eliminate non-competitive choices through the use of a Pareto front, and (c) provide summary tools from which the trade-offs between alternatives can be quantitatively evaluated and discussed. A structured but flexible process for contributing to team decisions is described for situations when all choices can easily be enumerated as well as when a search algorithm to explore a vast number of potential candidates is required. A collection of diverse examples ranging from model selection, through multiple response optimization, and designing an experiment illustrate the approach.

Keywords: multiple objectives, Pareto front, desirability functions, DMRCs

1. Introduction
We live and work in a highly interconnected world, where expertise and information from many sources is often relevant and beneficial to include when making important decisions. With the rise of discussion and implementation of Statistical Engineering, statisticians are embracing their role to “use known statistical principles and tools to solve high-impact problems for the benefit of mankind” (Hoerl and Snee, 2010). There are many promising opportunities to have a positive impact in the workplace when we participate in all aspects of process improvement and problem resolution (Anderson-Cook et al., 2012a and 2012b). However, there was little in my formal statistical training that prepared me for some of the roles that I am now advocating that we embrace as interdisciplinary team members to facilitate improved decision-making. My early practical training focused on giving a single strong justifiable answer that was anchored in incorporating uncertainty into the analysis of data. This statistical thinking emphasized that processes and systems have inherent uncertainty and that including this into our understanding was key to helping with solid defensible conclusions. What I have now come to realize is that there is another key contributor to good decision-making that we may underestimate and that may compromise the quality of our choices. To make problems manageable, we often seek to simplify them to make them well-defined and tidy. This can lead to (a) reducing the metrics of what is a good solution to too few dimensions, and/or (b) underutilizing the subject matter experts in the decision-making process.

When we consider only a single criterion, problems become much simpler and there is typically a single best answer. Beneath the simplicity lurks a danger of answering the wrong question and settling
for something that only partially meets our needs. Another way to paraphrase the point of Kimball (1957), is to say “There is no right answer to the wrong question.” Recently, my husband and I went shopping for a new car – we thought it would be straightforward as we went to the same dealership that had sold us our previous car. We test drove the latest model of the same car, and it seemed fine, except that we did not like the changes that had been made to the dashboard. Had it not been for this irritation, it would have been easy to choose this car as the easy answer. But driven by the desire for a different dashboard, we test drove a handful of cars. Looking at other cars, we were reminded how many different aspects there were that we should be evaluating. We were largely indifferent to options on some, but had strong preference on others. We discussed and debated the merits of various vehicles and what mattered to us, and when we finally chose the car there was a great deal of comfort with the decision that we had made. We knew the choice space of what was available, and we felt that we had intentionally chosen the best car for us. We also had improved understanding and empathy for the other’s priorities. In a microcosm, this is the opportunity that awaits when we explore the space of different criteria and different top alternatives more thoroughly.

The process of decision-making involves multiple stages, and having some organizational structure to the process can help guide the process as well as increase the likelihood of obtaining a reproducible result. In Anderson-Cook and Lu (2015), a 5-stage process, similar in spirit to the Six Sigma DMAIC (Hoerl and Snee, 2012, p 128-137) progression is described. Define-Measure-Reduce-Combine-Select (DMRCS) provides a structure for identifying and comparing alternatives based on multiple objectives. The Define step clarifies the scope of the optimization by identifying the problem to be solved, the appropriate characteristics to consider, and the appropriate set of solutions to consider. As with DMAIC, this step is critical for framing the key aspects of the decision to ensure that the correct problem is solved. It is important to make sure that all key facets of the decision are represented with an appropriate metric. It is important to think broadly about diverse aspects critical to the decision.

Another step that overlaps with the DMAIC process, is the Measure step, which focuses on the importance of having high quality data on which to base the decision. The quantitative criteria to be compared need to be precisely and accurately measured to allow fair and consistent comparisons between potential solutions. The Reduce step seeks to simplify and streamline in several dimensions. First, optimizing based on too many criteria often leads to mediocre results individually, as the trade-offs between criteria generally increase with more criteria. Hence, some triage of the priorities of the decision may lead to reduction in the number considered, as highly correlated responses or those measuring secondary objectives may be set aside until later in the process. Similar to designing an experiment where many potential factors are initially identified, before focusing on the most important ones, this part of the process allows due consideration of many potential choices to ensure the right subset have been selected. The second type of reduction to consider is eliminating non-contending solutions from further consideration. Constructing a Pareto front (discussed in greater detail later in the paper) is an objective way to achieve this.

The Combine step considers how to simultaneously look at information from different criteria, often measured on different scales and representing different facets of the decision. Here decision-maker priorities and how much they value good performance on different criteria is key to identifying the best solutions for the problem under consideration. Finally, the Select step identifies the available solution that is best suited to the decision-makers’ objectives and provides tools for making
comparisons between close contenders. At the conclusion of the process, the decision-maker should have an identified solution, and also be able to articulate why it is the best choice given the nature of the decision.

It should be clear that in many of these steps, there are opportunities for statisticians to participate in the process and contribute with expertise and guidance.

Consider four examples that we revisit in more detail at various points throughout in the paper:

**Example 1: Munitions Reliability Modeling:** To manage Department of Defense (DoD) stockpiles (populations) of munitions, it is important to be able to estimate their current and future reliability. Changes in reliability can be driven by multiple sources including aging, exposure to environmental conditions, and usage. To make decisions about which units to use or when units need to be removed from the stockpile, it is helpful to be able to make reliability predictions that are as specific to the experiences of individual units or small groups of units as possible. Statistical models can be used to identify different drivers of change in reliability and to generate predictions with an appropriate quantification of the uncertainty in that estimate (Anderson-Cook et al., 2015). Choosing which model best characterizes the reliability can have an important impact on management of the stockpile.

**Example 2: Stockpile Maintenance Prioritization:** In a related scenario, DoD managers may need to prioritize where to invest in the maintenance or surveillance of their stockpiles. In a given year, there may only be resources to dedicate to a subset of the stockpiles, and hence managers must balance different aspects of the decision (including estimated time until a problem, the consequence of a problem, and availability of alternative munitions) to decide which stockpiles receive investment. Different stakeholders assess the urgency of need based on differing perspectives.

**Example 3: Experiment Design Selection:** When designing an experiment, there are many considerations about what is an ideal design to meet the goals of the experiment. Myers et al. (p. 370, 2016) describe 11 aspects of a good design from which a subset of top priorities should be chosen to match the specific aims of the study. These aspects often compete against each other: For example, achieving better estimation of model parameter often comes at a price of protecting against model misspecification, and smaller less expensive experiment reduce the performance of many other metrics. Different specialized scenarios such as robust parameter design, restrictions on randomization or flexible cost constraints may each require different criteria.

**Example 4: Multiple Response Optimization:** When optimizing a process or product, there are often multiple attributes which need to be balanced, and simultaneously achieving the best possible performance on all of these is generally not attainable. If data were used to estimate models of the relationship between inputs and the characteristics of interest (for example, the responses of the experiment), then there is uncertainty in the resulting model parameter estimates, which should be actively incorporated into the decision-making process.

All of the examples share a number of characteristics which make decision-making more challenging: (a) there is considerable subject matter knowledge and expertise about the underlying process which can be leveraged, (b) there are several aspects to the decision which need to be simultaneously balanced, (c) there are multiple experts involved in selecting the path forward, and these...
decision-makers have different priorities and emphases that are often difficult to precisely quantify and historically might not even have been articulated.

This paper focuses on highlighting some areas where new perspectives, approaches and tools are needed to help statisticians expand their role when important decisions are being made. Section 2 looks at some of the things inhibiting some statisticians from fully participating in the decision-making process. Section 3 identifies some areas where statisticians can claim roles in the decision-making process. The examples are revisited to showcase how identifying and quantifying criteria to use in multiple objective optimization can lead to improved solutions (Section 4) and some tools that can be helpful for visualizing and comparing alternatives during the decision-making process (Section 5). Section 6 concludes with thoughts about the benefits of thinking more broadly about decision-making when explicitly considering multiple criteria.

2. What is holding us back?

If statisticians wish to expand into new roles, it is helpful to identify some of the obstacles to achieving this objective. First, much of our training has focused on univariate objectives and the subjectivity of accounting for different prioritizations of several objectives has been ignored or over-simplified. In the cases where multiple objectives are considered in the statistics literature, the simplifying assumption of all criteria being of equal importance is often asserted, with minimal justification. In my experience, this is rarely the case in practice. We often find ourselves unprepared to articulate how we make decisions that balance competing objectives, and so describing this process to others is awkward. The process of publishing research often forces simple storylines with the subjective choices about how to balance options seeming to be obvious and pre-determined.

Second, comparing several criteria requires deep understanding of the criteria themselves. If we are optimizing a design for a planned experiment, and we wish to estimate model parameters well, then D-optimality can be a good choice. When we optimize on a single criterion, it is not as important to have understanding about the relative size of several options, or be able to translate differences in values into tangible consequences. For a single criterion, there is a single best value and our goal is to find it. When we are in the realm of multiple criteria, then the relative size of different values of a given criterion needs to be interpreted and compared to the corresponding different on another criterion to see which option is more appealing. Hence with multiple criteria, there more pressure to understand what things mean, and also to be able to explain those differences to team members with less statistical training.

Third, I think that there has been a lack of available tools to help with the process of decision-making. Different people balance alternatives differently, and so it is unlikely that there is a single set of tools for handling this aspect. In Section 5 there is a description of some tools that we have develop and found to be helpful for visualizing and comparing alternatives. There is considerable opportunity to expand this set of tools and provide complementary tools that showcase differences between choices and highlight promising alternatives. Currently much of popular statistical software packages do not focus on giving alternatives, but prefer to offer a best solution given a precisely defined problem, instead of allowing users to consider multiple options. While there is a large set of problems that are well-handled by this approach, it would be helpful to have some available options for thoughtfully considering the multiple objective scenarios.
Fourth, many statisticians begin their careers with no formal training or experience with participating in an interactive process with subject matter experts from other disciplines. Presenting multiple alternatives and leading discussions about how to compare different choices based on several metrics is often unfamiliar and uncomfortable. If no formal process is used to choose between competing alternatives when a statistician makes decisions on their own, then guiding a discussion with experts who have dramatically different priorities is daunting. Facilitation of discussions and moderating conflict over different priorities requires experience and leadership skills.

Finally, leading the group in negotiations for which solution is best for difficult problems demands credibility from the subject matter experts participating in the process. Many statisticians are relative newcomers to leadership roles on interdisciplinary teams, and so earning this role requires experience and respect often based on successes in previous projects.

3. Defining the statisticians role in decision-making

At the 2015 American Statistical Association Joint Statistical Meetings in Seattle, the President’s Invited Address was given by Christine Fox, who served as the acting United States Deputy Secretary of Defense and Director of Cost Assessment and Program Evaluation in the Office of the Secretary of Defense. She articulated her vision of some key aspects that decision-makers want from statisticians and analysts. She emphasized that decision-makers are hungry for tools to make better decisions, and that bigger decisions often mean harder choices (with more difficult trade-offs). She encouraged statisticians to participate in the process as facilitators and helpers, not as gate-keepers throwing in obstacles to progress. She also said that it was critical that statisticians ensure that the solutions that they offer match the actual problem to be solved. An important aspect to remember is that more senior decision-makers make more decisions and have less time for each one. Hence, efficient communication of option with all essential information included and explained is key to making valuable contributions. Based on this framework, here are a few areas that I propose where statisticians can expand their skills to be able to better participate in the decision-making process.

First, an awareness of different roles for statisticians in the decision-making process can help with assessing what the team is expecting. There are a range of levels of involvement that are possible: at the lowest level of involvement might be tasks that fit into the category of “decision support”. In this capacity the statistician might be asked to perform analyses or collect data that are used as part of a broader collection of information to guide the decision. Here technical expertise is being sought, but an opinion or contribution to the decision is not anticipated. At the other end of the spectrum is “provide the decision”, where the statistician is expected to provide a final answer with justification. Both of these roles have been thoroughly explored in the literature and with numerous examples. A middle category, “decision analysis”, has the statistician providing decision support as well as being heavily involved in the decision-making process, but without the exclusive final say in making the selection. The team is expecting statistician participation, advice and recommendations that are combined with knowledge and opinions from other subject matter experts for a team consensus on how to proceed. Different situations require different levels of involvement, and the stakes of the decision often guide who is involved and the degree of responsibility that each has.

Second, similar to the statistical thinking revolution several decades ago that highlighted the importance of thinking about systems as interconnected processes with variation, statisticians can play a
key role in helping to frame problems as the balancing of competing objectives. Very few problems that do not have obvious solutions can be summarized by focusing on a single objective. A stark reminder of this came when I served on a National Academy of Science panel for evaluating testing protocols for combat helmets (Nair et al., 2014). At the first meeting, one of my fellow panel members stood up and said “we can build a helmet that no bullet will penetrate.” There was a long pause while we thought “so, what are we doing here?” He then continued and said “the problem is that it would be so heavy that no soldier could move when they were wearing it.” So this sharpened our focus – the goal was to think about realistic standards of safety for the soldiers, while still maintain sensible weight constraints that allowed them to move, elude attack and perform their duties. Most challenging problems involve balancing between competing criteria. Statisticians can help identify when problems have been oversimplified, and what aspects are not being actively considered.

Third, redefining the process of analysis and drawing conclusions to integrate the subject matter experts throughout can be beneficial. I see the statistician’s role more broadly as being included in a process that iterates with the subject matters throughout and concludes with the presentation of multiple alternatives with clear articulation of advantages and disadvantages for each potential choice. In many ways the objective presentation of alternatives is not something for which many of us have been formally trained, and it also may be unfamiliar ground for our own personal decision-making. The disciplined process of looking at multiple alternatives and articulating their merits is not something that many people practice. Think about the last time you made a major decision, like buying a new car or looking for a job. Was there a formal process of assessing the merits of the options, or did you use a “gut-feeling” approach for the decision? Indeed, research (Carlson et al., 2006 and Blanchard et al., 2014) has shown that few of us use formal methods for evaluating choices and whichever option seems initially appealing tends to maintain disproportionate favor throughout the decision-making process. Marketers exploit these tendencies to lure us in with a first glimpse of an option and then hope that our decision-making process will be irrational or inconsistent enough to allow that choice to stand.

Finally, when we recognize that there are many right potential solutions for a given problem, we are liberated to think about what distinguishes them and what makes one choice better than another in different circumstances. This means that we have to stop thinking that “subjective” as a dirty word. When we have eliminated the irrational choices from consideration (why would you ever choose a more expensive lower quality solution if alternatives existed?), then what distinguishes the remaining choices is how we value the trade-offs between the different aspects of the decision. To make good decisions, we need tools to allow us to explore alternatives, understand their strengths and weaknesses, and combine that with our ability to prioritize the criteria for the best answer for our problem. With different experts involved in the decision-making process, there is also a need for these tools to facilitate discussion between people with conflicting priorities to reach a solution that is amenable to all.

4. Identifying what criteria to consider

In the collaborative environment at Los Alamos National Laboratory, we often have multiple statisticians working on different phases of a project. Over the years, I have been amazed at how differently we think about the same problem given our historical experiences. For example, when designing an experiment, it would be rare that each statistician would not initially identify a slightly (or dramatically) different solution given the same information from talking with subject matter experts.
This should be a cautionary reminder for us that we all have filters that make subjective choices about how to begin to solve a problem. From different starting points, the solutions can diverge and lead to substantial differences in the final result. I think that my experience at LANL has greatly improved my ability to articulate my priorities, listen and assimilate differing valid views, and to realize that there are many correct perspectives about how to think about the same problem. The quality of our solutions can be dramatically improved if we are able to leverage the collective wisdom of experts to incorporate their historical experiences.

Consider the Stockpile Maintenance Prioritization (example 2 in Section 1) where the goal is to identify several stockpiles in need of investment to maintain their availability. In the initial meeting of experts, the leader of the committee charged with making these choices started with the statement that the task was quite simple: the team would look at the reliability and predict when each stockpile would cross a pre-determined threshold and become unacceptable. Using this criterion, a ranking would be established. As a statistician, I was asked to help with determining how the reliability should be predicted. Several members of the group looked unhappy and seemed to disengage within minutes. One member protested and suggested that what reserves of units were available should be considered. After some animated discussion about which metric was better, I suggested that we itemize many possible metrics and that there were methods to consider more than one priority and that how we valued different aspects could be evaluated later. This changed the focus of the meeting, and led to a list of more than 20 possible options for how we might evaluate how urgently a stockpile needed attention. After the list was compiled with vigorous participation from the entire group, we returned in a subsequent meeting to examine the list and place the choices into common groups. It emerged that there was several area: One focused on the reliability of the units (how soon would the mean or some percentile of the population reach a critical threshold, the rate of change of reliability, reliability at some current or future time, etc). Another group considered availability (reserves relative to current usage, additions through acquisition, alternative stockpiles that would provide acceptable substitutes, etc.). Other groups focused on the consequences of a shortage of units, and logistical constraints of making the units available for service. Once this grouped list of options was available to the team, the next stage involved determining what data were available to quantify these criteria and how they could be translated into a precise qualitative metric. For some criteria, there was insufficient data available. For others, the qualitative information could be converted to an expert score with carefully defined criteria. For other criteria, a quantitative metric could be calculated or estimated. After there was clarity on available metrics for the different aspects of the problems, it was more straightforward to look for duplicate summaries, gaps in information that was deemed important to the decision-making process, and then select a primary and secondary set of metrics on which to focus. This process changed the course of the resource allocation exercise: It encouraged engagement of all members of the team, since multiple criteria could be discussed and considered. It encouraged broader interpretation of the goal of the study with more realistic assessment of what a successful outcome would be. It allowed evaluation of the impact of multiple aspects of the stockpiles to be considered and compared when identifying candidates for additional resources.

In Multiple Response Optimization (example 4) scenarios, there are often many measurements that are taken when an experiment is run. Some are the focus of the study, while others might be collected simply because it is straightforward and inexpensive to gather that data. In these cases, it is
important to think about how to divide the responses into primary and secondary priorities, since trying to optimize across more than a moderate number of alternatives can often lead to overall mediocre performance. In addition there may be several metrics that summarize similar characteristics of the process being measured. In these cases, it is helpful to think about data quality and select continuous responses over categorical ones when possible.

The Experiment Design Selection (example 3) scenario is often complex. Myers et al. (p. 370, 2016) describe 11 aspects of a good design, which can be broken into several broad categories. A primary consideration is often to estimate model parameters or predict new observations well. This, of course, is contingent of having selected an assumed model, which is hoped to be capable of summarizing the true but typically unknown relationship between the chosen inputs and the responses of interest. Historically, if computer-generated optimization is being used, then D-, G- and I-optimality (Myers et al., 2016) have been common choices. A second category of criteria could consider evaluation of model assumptions such as homogeneity of variance or the choice of model. Next we may wish to evaluate how the design would perform in the event that something does not go as planned. Other aspects can also include flexibility of implementation or augmentation. Finally, cost (often measured by the size of the experiment) is a consideration that is important in most experiments. In sequential experimentation, it is helpful to think about conserving budget to have it available in subsequent stages of the data collection. Many categories conflict with each other, resulting in trade-offs when choosing an appropriate design. For example, good choices for designs when the model selected is correct differ considerably from cases where we are trying to evaluate if the model is correct. Another potential trade-off considers balancing if the experiment will run smoothly versus if there might be some complications. Finally, design size (cost) trades off with most of the other categories: Larger designs improve performance for most of the other categories, but use more resources.

There are many examples in the literature that illustrate different choices of criteria: Lu, Anderson-Cook, Robinson (2011) use D-efficiency and two other criteria to balance good estimation when the assumed model is correct with protection if the assumed model is not adequate. Lu and Anderson-Cook (2012) consider when good estimation is traded-off with evaluation of lack of fit and good estimation of pure error. Lu and Anderson-Cook (2014) and Lu, Robinson, Anderson-Cook (2014c) consider good estimation and cost considerations for split-plot designs where the relative cost of the whole plots and subplots vary. Lu, Anderson-Cook, Robinson (2011) discuss a robust parameter design scenario where good estimation of both the mean and variance models is important. Lu, Johnson and Anderson-Cook (2014b) select good screening designs from a catalog of regular and non-regular designs using criteria commonly associated with supersaturated designs. Lu, Li, Anderson-Cook (2015) balance consumer’s and producer’s risk with the probability of passing the test and test size for designing a demonstration test. Lu, Chapman, Anderson-Cook (2013) consider good estimation of system and sub-system reliability when evaluating which of several different data types with different associated costs to collect from a complex system. In each of these cases, the criteria were carefully matched to the goals of the study and the experts’ confidence in the assumed model.

Throughout my career, I have encountered many experimenters who have been given an “optimal design”, but when questioned about it, they were at a loss to say over which criterion it was optimized. First, I believe that optimizing over a single objective is almost always too simplistic a solution for most design of experiment scenarios, and second, too often “standard choices” are made about
what the right criterion should be and the assumed model on which it is based. As shown in the next section, there are tools available that allow a tailored choice of criteria to be made that provide a good match for the important criteria for each particular study. Anderson-Cook (2013) discusses some of the thought process for choosing between criteria and choosing the right number of criteria over which to formally optimize.

5. Tools for identifying contenders and comparing alternatives

In the process for decision-making involving multiple criteria, there are two areas where tools can be highly beneficial to the process: The first involves having an efficient way of objectively eliminating non-contenders from further consideration, while allowing the decision-maker to see what viable choices are available. The second focuses on strategies and visualization tools for subjectively choosing between the contenders that best match the specific priorities of the study.

5.1 The Pareto Front for Identifying Contenders

A key component of multiple criterion optimization is to separate the idea of objectivity and subjectivity, and be aware of choices that introduce subjectivity into the decision-making process. In this first stage of the process, the goal is to remove irrational candidate solutions from further consideration. This fits in nicely as part of the Reduce stage of DMRCS, where after making strategic choices about which criteria are important in the decision, the decision-maker wants to remove candidates that should not be considered viable. For example, if the decision involves a trade-off of minimizing cost versus maximizing quality, then we should never consider (i) a higher priced option that is lower quality, (ii) a higher priced option with the same quality, or (iii) an equivalently priced option with lower quality. Hence, we can define the Pareto set as the set of candidates that are non-dominated by any other solution. A dominated solution is one where there is at least one other solution that has criterion values at least as good as the dominated solution for all criteria, and strictly better for at least one criterion. Sun et al. (1997) used the terminology of “admissibility” to match the non-dominated definition in the Pareto front literature. The Pareto front is the set of criteria values for the candidates in the Pareto set. The Pareto front gives the range of criterion values of rational choices from among which the decision-maker should select a final solution, conditional on there being agreement on the criteria being used. The DMRCS process facilitates first deciding what criteria are important, and then later looking at the particular solutions in this context. Separating the choice of criteria from identifying which solutions look best under a set of criteria helps avoid the post-hoc rationalization of “a favorite solution” by selecting criteria to make it a legitimate choice.

For example, consider trying to select a best option based on maximizing two criteria, C1 and C2, from among the 20 solutions shown in Table 1. Figure 1 shows the plotted responses, with the ideal solution being located in the top right corner of the plot. As is typical, none of the 20 solutions, simultaneously maximize both C1 and C2, and hence there are trade-offs needed between the two criteria. When we construct the Pareto front, we find that there are only 4 candidates (shown with solid circles in Figure 1 and in bold in Table 1) that are not dominated by others. For all of the other 16 candidates, there is at least one of the solutions on the Pareto front that dominates it. For example, candidate 4 (4,0.6) is dominated by 3 (4,0.9), 10 (6,0.7) and 19 (9,0.6). Candidate 15 (8,0.4) is dominated by 19. By constructing the Pareto front, we have eliminated 16 of the candidates from further
consideration, and have full reassurance that unless we change the criteria used for the decision, the eliminated candidates are not viable.

The definition of the Pareto front and constructing it from a list of candidates remains straightforward for any number of criteria using software available in JMP (2014) or R (available from the author by request).

Table 1: Candidate solutions

<table>
<thead>
<tr>
<th>#</th>
<th>C1</th>
<th>C2</th>
<th>#</th>
<th>C1</th>
<th>C2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>0.7</td>
<td>11</td>
<td>7</td>
<td>0.3</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>0.8</td>
<td>12</td>
<td>7</td>
<td>0.2</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>0.9</td>
<td>13</td>
<td>7</td>
<td>0.1</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>0.6</td>
<td>14</td>
<td>7</td>
<td>0.5</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>0.3</td>
<td>15</td>
<td>8</td>
<td>0.4</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>0.5</td>
<td>16</td>
<td>8</td>
<td>0.2</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>0.3</td>
<td>17</td>
<td>8</td>
<td>0.3</td>
</tr>
<tr>
<td>8</td>
<td>6</td>
<td>0.6</td>
<td>18</td>
<td>9</td>
<td>0.1</td>
</tr>
<tr>
<td>9</td>
<td>6</td>
<td>0.4</td>
<td>19</td>
<td>9</td>
<td>0.6</td>
</tr>
<tr>
<td>10</td>
<td>6</td>
<td>0.7</td>
<td>20</td>
<td>10</td>
<td>0.2</td>
</tr>
</tbody>
</table>

When there is no fixed set of candidate solutions readily available, then populating the Pareto front requires additional effort. In the case of example 4 (Multiple Response Estimation) it is possible to evaluate the estimated response surface equations at all locations of an identified grid of input combinations across the region of operability. The finest or coarseness of the grid should be based on practical considerations of what precision there is in setting the input factors. See Chapman et al. (2014a) for more details and an example. Since the estimated response surfaces have associated uncertainty, the Pareto front will change based on parameter estimates. Chapman et al. (2014a, 2014b, 2015) discuss a variety of strategies for incorporating this uncertainty into the Pareto front summaries and consider its impact.

For Experiment Design Selection (example 3), a search algorithm can be highly beneficial as the possible number of designs becomes very large for even a moderate sized design scenarios. Lu and Anderson-Cook (2013) describe the PAPE (Pareto Aggregating Point Exchange) algorithm that can find and add candidate designs to the Pareto front for flexible design criteria. This algorithm assumes that the user has a specified design region with a set of candidate points from which to select. What distinguishes the algorithm from standard point exchange approaches based on a single criterion is (1) the choice of how to determine when to update current solution with a new improved option, and (2) every candidate design that is create is considered for inclusion on the Pareto front.

Expanding point (1), when there is only a single criterion, it is straightforward to determine if a new solution is an improvement over the current one. However, when there are multiple criteria, it is less obvious how to proceed. The PAPE algorithm uses a desirability function approach for combining the criteria into a single score to determine when an improved solution has been found. The search algorithm uses a variety of weights spread throughout the set of all possible weights to encourage populating the entire Pareto front. Lu, Anderson-Cook and Lin (2014a) suggest some modifications to
streamline the search process if only a portion of the possible desirability function weights are of interest.

The second modification to the algorithm (2) means that as the point exchange algorithm is implemented, each new candidate design is considered for inclusion on the Pareto front. Frequently a time-consuming part of the algorithm is evaluating the criteria for each candidate, and once this is done there are little additional resources required to determine if it is dominated by current solutions or not. Hence for each random start of the PAPE algorithm, multiple solutions can be added to the Pareto front, which can greatly speed the search process.

The algorithm can be adapted depending on whether there are secondary criteria that will be considered when making the final decision. If only the criteria that are used in the search algorithm are to be used, then ties (those solutions with exactly the same values for all criteria) can be ignored with only one candidate for each set of values retained. If there are secondary criteria are also being considered, then it is important to have provisions in the search algorithm to retain all solutions with the same primary criteria values as they may differ on the secondary criteria.

The Pareto front is the natural starting point for decision-making for multiple criterion optimization, because it has several advantages:

(a) Pareto fronts can be constructed for any number of quantitative criteria.
(b) Non-contenders can be removed from discussion, since undisputed alternatives exist that are superior for the identified criteria.
(c) The construction of the Pareto front is completely objective.
(d) The Pareto front reduces the number of solutions that needs to be considered, making the problem more manageable.
(e) The range of values for each criterion from the Pareto front can help shape discussion about the degree of trade-off between the different alternatives, before starting to consider individual solutions.

In the next section, we consider different alternatives about how to evaluate contenders on the Pareto front based on the priorities of the decision.

5.2 Strategies for Subjectively Choosing a Best Solution

After the initial objective stage of constructing the Pareto front, a reduced number of candidates are available for more detailed consideration in the decision-making process. The main difficulties in evaluating which solution on the Pareto front is best for a particular study are (1) the criteria are often measured on different scales and hence are difficult to compare directly, (2) different decision-makers attach different priority to the individual criteria, and (3) decision-makers may follow different paths to their decision-making. To accommodate these challenges, Lu, Anderson-Cook and Robinson (2011) proposed clear articulation of the subjective choices made to increase the transparency of the process, and using a variety of graphical summaries that allow exploration of different decision-making strategies and priorities. Myer, Montgomery, Anderson-Cook (Ch. 7, 2016) describe some of the tools available. Depending on the dimension of the problem (i.e. the number of criteria under consideration), the tools may need to be adapted.

Since decision-makers can follow different thought processes when comparing potential solutions, it is helpful to provide several tools for visualizing the data. Initially it can be helpful to look at
pairwise scatterplots of the Pareto front (similar to Figure 1), with a plot for each pair of criteria. This is helpful for evaluating the degree of trade-off between the criteria, as this can differ for various pairs of criteria. For some pairs it may be possible to nearly achieve the optimal value for both criteria simultaneously, while for other harsh sacrifices may be needed from one criterion to improve the other. Understanding these relationships can help guide decision-making and also help multiple decision-makers realize that all stakeholders may not be able to have everything that they want. One potential approach, called the threshold approach, orders the criteria by importance and sequentially imposes acceptable thresholds for each. For the simple example in Figure 1, we might say that we want to maximize C1 conditional on C2 being at least 0.5. For this set of choices, the best solution would be candidate 19 (9,0.6). Alternately, we may want to maximize C2 conditional on C1 being at least 6, leading to candidate 10 (6,0.7). If more than 2 criteria are considered, then constraints are needed for all but one of the criteria. Note that whether we consider the complete set of data, or just the Pareto front, the identified solutions from the threshold approach are always located on the Pareto front. This approach is appealing in that it uses the data in its original units and hence the natural interpretation of the criteria is preserved. Disadvantage of this approach include that it is quite dependent on the order that the criteria are considered, and when there are multiple criteria it is quite common to end up with a null set of choices if the constraints considered are too ambitious. In discussing this with different decision-makers, there appears to be little agreement about whether to start or finish with the most important criterion. It can be helpful to experiment with several different orders of criteria thresholding to evaluate the robustness of the selection process.

An alternative to the threshold approach utilizes desirability functions (Derringer and Suich, 1980) to combine the criteria into a single desirability function score. The first stage involves translating each criterion onto a common scale (frequently [0,1], where 1 is most desirable and 0 is least desirable). If done thoughtfully, similar desirability scores for different criteria can be thought of as comparable. Next, a functional form is chosen to combine the scaled criteria into a single score for each of the candidates on the Pareto front. Common choices include additive \( DF_j = \sum w_i C_{ij} \) and multiplicative forms \( DF_j = \prod C_{ij}^{-w_i} \), where \( i \) indicates the \( i \)th criterion and \( j \) indicates the \( j \)th candidate solution. Finally, the weights \( (w_i) \) are chosen to quantify how much to emphasize each criterion. There are several subjective aspects to this approach, which give reason to use it thoughtfully and with awareness of where subjectivity can have an impact. The translation from original units to the desirability scale requires care in that the best and worst values from the Pareto front might span “excellent to terrible” for one criterion, but “excellent to moderate” for another. If not done thoughtfully, then the induced desirability value of 0 can have very different impacts on the final choice. The additive combination of individual scores is more tolerant of poor results from any particular criterion as a very high score from other criteria might overcome this disadvantage. The multiplicative form penalizes a zero or near zero desirability score more severely. The weights in the desirability score are often difficult to precisely quantify and might be better considered as ranges. All of these subjective elements can be varied with repeated analyses to assess the robustness of the decision to these choices. An advantage of this approach is that it encourages the upfront choice of these subjective pieces, before specific solutions are identified. This can encourage more straightforward discussion among decision-makers with
different priorities. A disadvantage of the approach is that it disguises the original criteria values, and if not performed thoughtfully can lead to a “black-box” feel to the solutions.

Once the desirability function structure has been selected, then there are a number of graphical summaries that can be used to highlight some of the top choices as well as their strengths and weaknesses. These plots can facilitate discussion among decision-makers and showcase the trade-offs between alternatives. It is also possible to combine these plots with scatterplots of the original values to remain grounded with their original scales. Some of the plots to consider include:

a. Trade-off plot – a single overlaid line plot with one line for each criterion that shows their desirability scores for all top solutions (Lu et al., 2011)
b. Mixture plot – a single plot showing which solution is best for all weight combinations of the desirability function (Lu et al., 2011)
c. Synthesized efficiency plot – one plot for each potential solution that summarizes how well it performs relative to the best available option at any particular weight combination of the desirability function (Lu and Anderson-Cook, 2012)
d. Fraction of Weight Space (FWS) plot – a single plot with a line for each potential solution which summarizes the synthesized efficiency values across the range of weights of interest (Lu et al., 2013 and Lu et al., 2014a)
e. Input Grid plot – a single plot of input combinations for the multiple response optimization scenario that highlights which locations of the design space appear on the Pareto front (Chapman et al., 2014a and 2014b)
f. Desirability-Weight-Input-Volume (DWIV) plot – a set of linked plots for the multiple response optimization scenario that highlights the top solutions’ desirability scores (similar to a trade-off plot), range of weights for which it is best, input factor values and fraction of volume for which the solution is optimal (Lu et al., 2016). This plot is particularly helpful when the dimensionality of the problem becomes larger.

We return now to the examples to highlight some of the choices that were made about how to facilitate the decision-making process. For the Munition Reliability Modeling example (example 1), it is possible to consider several competing model selection metrics (such as AIC, BIC, DIC, Median Posterior Model or Prediction-Based criteria) to identify a top subset of contending models (Anderson-Cook et al., 2015). In this case, the role of expert opinion is particularly important to incorporate since there were multiple models that were very close in performance based on the metrics. Plots of reliability as a function of age and the other explanatory variables allowed the subject matter experts to see the estimated reliability patterns. This showcased the relationship between potential explanatory factors and reliability. Based on current scientific and engineering understanding, the experts were able to identify top models that best matched expected patterns as well as gain understanding about newly revealed patterns. Here the screening process of using model selection metrics was used to identify a manageable number of models to investigate further with the experts.

The Stockpile Maintenance Prioritization example (example 2) sought to identify the top N stockpiles that most urgently needed monitoring and maintenance resources. Burke et al. (2016) modified some of standard Pareto front graphical summaries to look at layered Pareto fronts. This enabled the subject matter experts to see which stockpiles were identified as most urgent for any set of
desirability function weight combinations of reliability, availability and consequence. The elimination of a large fraction of the stockpiles from consideration, based on them never being in the top N for any weight combination helped to streamline the discussions. The multiple experts involved in the decision-making had widely differing priorities, but the graphical summaries helped to facilitate discussion and guide a consensus set of choices.

At LANL, we are frequently involved in designing important high consequence experiments. One of the benefits that we have realized with the multiple criteria optimization approach has been improved engagement in the design selection process. With it comes an increased responsibility for the statisticians to provide accessible descriptions of different possible metrics and how to interpret the values of these metrics, but the discussions about which metrics to use has often led to improved understanding of study goals and priorities. The ability to adapt the Pareto front search methods for different design metrics has been important, because as noted in the previous section, a wide variety have been used in different experimental scenarios. Once a smaller subset of potential designs have been identified, we have been pleased with the detailed discussions that we have had with the decision-makers about which choices suit their needs best. In many cases, top contenders were compared using secondary criteria to help guide the final choice. In addition, we have noted increased feeling of ownership about the final design and greater understanding and appreciation that the right design has been selected.

Example 4 (Multiple Response Optimization) is a very common scenario resulting from designed experiments. Given the economic and time resources needed to run experiments, it is typical to collect data on multiple responses. Traditional approaches such as overlaying contour plots or using constrained optimization have several shortcomings: First, they are difficult to implement as the dimension of the input space and/or the number of responses increases. Second, they aim to provide a single best location in the design space, rather than providing alternatives that allow for understanding of different regions of good performance. Perhaps the most important weakness of some traditional approaches is their inability to incorporate the response uncertainty into the decision-making process. The Pareto front approach suggested in Chapman et al. (2014a and 2014b) provides potential strategies for considering this uncertainty directly. Chapman et al. (2014a) considers both the mean model based on the point estimates of the response surface model parameters, as well as a worst case predicted value. The Pareto fronts and suggested input combinations can then be compared to check for robustness as well as providing more realistic summaries about what performance is likely once a given solution is implemented. Chapman et al. (2014b) use simulations based on the distributions of the estimated model parameters to provide summaries of differences in the Pareto fronts identifies as well as the range of response values. This exploration allows for greater understanding and reveals how different amounts of variation in the responses can impact robustness.

There are many other scenarios where the formal structure of the DMRCS process can help to guide decision-makers as they consider how to balance multiple competing objectives. The framework is general enough to allow great flexibility and accommodate differences in priorities.

6. Conclusions

Playing a larger role in the decision-making process provides challenges for statisticians, but there are potentially substantial rewards, both for us but also for all those involved in the decision. First,
when we engage in identifying the right criteria to focus on, and have good supporting tools to identify and compare the multiple alternatives that match those criteria, we have the opportunity to make a better decision. With a systematic approach, there can be consensus about what aspects are important, and understanding of what inherent trade-offs are necessary when considering a chosen collection of goals. By examining multiple contenders, we gain a deeper understanding of what is available, and how neighboring solutions are related to each other. With this understanding, we are able to carefully consider what is important, what trade-offs we are willing to make and what is the best choice for our priorities. As a result, we are likely to end up with a better solution and also have deeper appreciation for why it is the right choice. By having participated in a process where active engagement and weighing of alternatives was required, the decision-maker has greater buy-in on the final choice and is prepared to defend the choice in a much more compelling way.

When multiple decision-makers are involved in the selection of a final choice, the discussion facilitated by having multiple alternative to compare based on quantitative metrics can lead to improved opportunity for consensus, when it becomes apparent that everyone cannot get what they want. In addition, since group dynamics are guided by a formalized process, there is a greater opportunity that the most persuasive person does not get their choice, and that the decision remains data-driven.

I think there are still many opportunities for additional quantitative and graphical tools to be developed to suit different decision-making styles and preferences (Anderson-Cook, 2015). This is an area where statisticians have much to offer, and we should not be deterred by the subjective aspects of the process, as participation in the decision-making process can elevate the prestige of statisticians and allow them to inject quantitative data-driven aspects in an essential part of the business process.

References

Quality Engineering 26 269–284.