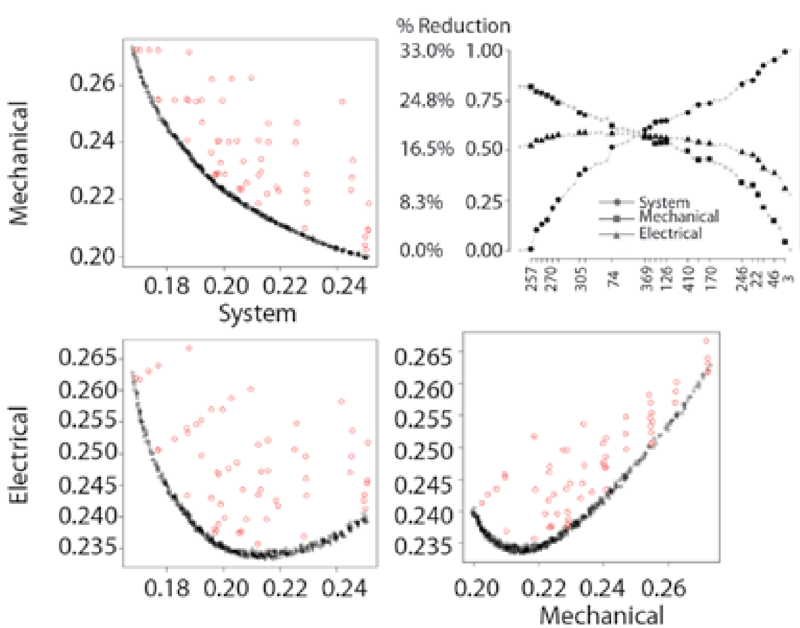
Strategic Data Collection for Optimal Use of Resources Based on Multiple Criteria

Christine Anderson-Cook,
CCS-6

The LANL Statistical Sciences Group, CCS-6, has an established record of research in experiment design and resource allocation. The goal of these efforts is to provide a quantitatively justified approach for determining what future data should be collected based on specific programmatic goals. The Pareto front approach allows decision-makers to establish specific objectives on which to focus for future data collection and, to identify superior choices to consider further, and then provide graphical summaries to compare alternatives with the goal of identifying a best choice. The approach allows flexibility for how the different objectives are weighted in the optimization, while allowing examination of the robustness across different prioritizations. This article describes a simplified example of how the methods can be used to maximally reduce uncertainty associated with system and subsystem reliability estimates.

As budgets become more tightly constrained, making thoughtful and justifiable decisions about what data to collect becomes increasingly important. In addition, there may be several types of data with different associated characteristics, costs, and precision from which to select. The Pareto front methodology adapted to experiment design and data collection [1] provides an approach for making quantitatively based decisions that can be tailored to the particular needs of a study. The process involves several steps. Step 1 requires that one or more top objectives for the study be identified and a quantitative metric determined that summarizes the performance based on each objective. There is considerable flexibility to choose measures that represent different aspects of a good solution and are specific to the current study. Step 2 is an objective step where the Pareto front [2] is constructed (by selecting superior candidates either from a list of potential solutions or performing a formal optimization search). Potential solutions not on the front can be discarded from future consideration since they are strictly inferior to at least one solution on the front and hence do not represent rational choices. Step 3 examines solutions on the front and allows examination of their performance on each of the criteria separately. In addition, the decision-maker can examine robustness to different desirability function weights [3], trade-offs between criteria, and overall performance relative to the best available for different weights [4]. Based on the specific needs of the study and understanding the alternatives, the decision-maker can select which data should be collected.

Fig. 1. Pairwise scatterplot and trade-off plot of Pareto front for system and subsystem reliability estimation example.



To illustrate the methodology, consider a series system composed of two subsystems (mechanical and electrical), with two and four components, respectively. There are nine different types of pass/fail data available (1 system + 2 subsystem + 6 component), each with different associated costs. (In previous years, some data were collected from each of these data types, and a Bayesian analysis [5,6] was used to estimate the reliability of the system and all of its parts). The primary goal of collecting new data from a fixed budget is to maximally improve the precision of the system reliability estimate. This can be quantified with the width of the 95% credible interval from the analysis. In addition to this objective, it is also desirable to improve the precision of the estimates for the two subsystems (as measured by the widths of their credible intervals).

Figure 1 shows a pairwise scatterplot of the Pareto front for these three objectives, along with a trade-off plot [1] to capture the relationship between objectives. In the scatterplots, the red circles represent proposed allocations of the resources, which were used to seed the search for best allocations using a genetic algorithm. The black circles represent solutions on the front, where ideal solutions would be located in the bottom left corner of each plot (minimizing the credible interval

For more information contact Christine Anderson-Cook at candcook@lanl.gov.

Funding Acknowledgments
DoD/DOE Joint Munitions Program and Advanced Certification

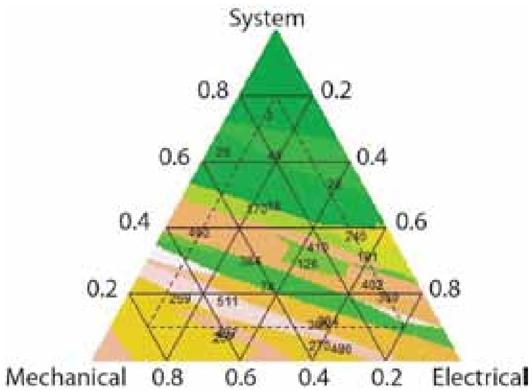


Fig. 2. Mixture plot of best allocations for different weight combinations of reductions for the system and subsystem credible interval widths.

widths). As is typical, no new collection of data is optimal for maximally reducing the width of all three credible intervals simultaneously. Hence, trade-offs between alternative solutions now are considered to arrive at a decision. In the top right corner of Fig. 1, different solutions are shown with the associated reduction in width from the current estimates. At the far right of the figure, we see a solution for this scenario that maximizes improvement (approximately a 33% reduction) for the system, but performs relatively poorly for the two subsystems. Alternatively, in the middle of the plot, there are some solutions that achieve at least a 16% reduction in width for all three criteria.

Figure 2 shows a mixture plot of the allocations that are best for particular weight combinations of the criteria using the additive desirability function [3] scaling of objectives onto a common scale. The vertices correspond to placement of all the weight on a single criterion, the edge represents the combination of just two objectives with non-zero weights, and the interior corresponds to non-zero weights for all three criteria. Note how allocation 3 labeled and shaded in green at the top of Fig. 2, is the near-best for maximal reduction of the width of the system reliability credible interval, corresponds to a moderately large region near the System (top) vertex. Different alternatives are suggested as best depending on the decision-maker's valuation of the different contributions.

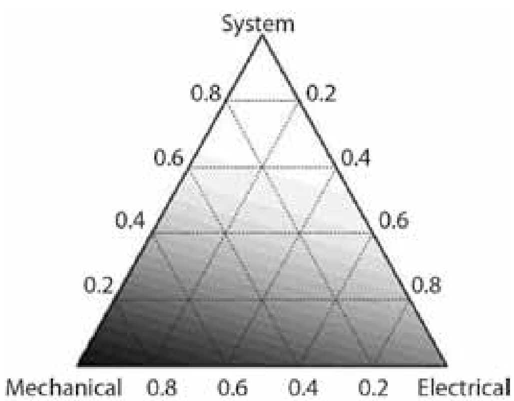


Fig. 3. Synthesized efficiency plot for a particular allocation to compare its performance against the best possible allocation for various weight combinations.

Finally, Fig. 3 shows the synthesized efficiency [4] of allocation 3 relative to the best design possible for each weight combination of the three criteria. This allows the decision-maker to determine the adequacy of its performance for the subsystem to objectives. For more details on the analysis and decision-making process for this example, see [7]. The suite of graphical methods helps describe the interrelationship between criteria and the possible performance across all of the objectives when considered together.

The advantages of the Pareto front approach include its flexibility for accommodating any possible quantitative measures for characterizing the goals of a data-collection study. This flexibility allows the decision-maker to solve the right problem. In other scenarios, we considered combining physical and first principles modeling data and optimizing based on good prediction, estimating tuning parameters and calibrating the data through a discrepancy function. In addition, by identifying the entire Pareto front of potential candidate solutions, the range of possible values for each criterion is available to calibrate comparisons. This method can be applied to a wide variety of data collection problems that allow leveraging current understanding of the problem and quantifying the anticipated benefit the additional data will provide in advance.

[1] Lu, L. et al., *Technometrics* **53** 353 (2011).

[2] Kasprzak, E.M. and K.E. Lewis, *Structural and Multidisciplinary Optimization* **22**, 208 (2001).

[3] Derringer, G. and R. Suich, *J Qual Tech* **12**, 214 (1980).

[4] Lu, L. and C.M. Anderson-Cook, *Qual Eng* **24** 404 (2012).

[5] Anderson-Cook, C.M. et al., *J. of the Military Operations Research Society* **13** 5-18 (2008).

[6] Chapman, J. et al., *Technometrics* **54** 410-424 (2012).

[7] Lu, L. et al., "A Case Study on Selecting A Best Allocation of New Data for Improving the Estimation Precision of System and Sub-System Reliability using Pareto Fronts," LA-UR 12-21268 (2012).