LA-UR- 12-00611

Approved for public release; distribution is unlimited.

Title: Strategic Future Data Collection for Complex Systems Based on Multiple Reliability Objectives

Author(s): C.M. Anderson-Cook
L. Lu
J. Chapman

Intended for: Conference on Data Analysis
Santa Fe, NM
March 1, 2012

Los Alamos
NATIONAL LABORATORY
EST. 1943

Strategic Future Data Collection for Complex Systems Based on Multiple Reliability Objectives

Christine Anderson-Cook, Statistical Sciences Group, Los Alamos National Laboratory

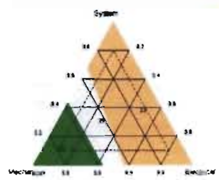
Lu Lu, Statistical Sciences Group, Los Alamos National Laboratory

Jessica Chapman, Math, Computer Science and Statistics Department, St. Lawrence University

The talk presents an example of multiple criterion optimization in the context of sequential experimentation. When estimating the reliability of a complex system comprised of many components, there are often different types of data which might be collected at the system, sub-system and component levels. We consider an example where the goal is to better estimate the reliability of a system and two of its major sub-systems. After initial analysis has been performed, there is an opportunity to collect more data. The goal is to select new data which maximally improves the quality of the estimation of the system and sub-system reliabilities. The talk presents a process with accompanying graphical tools based on the Pareto front approach to multiple criterion optimization, which allows some possible collections of new data to be eliminated as clearly inferior. From the remaining allocations, the best set of new data can be identified based on the relative importance of the different criteria, and the anticipated improvement in the quality of prediction quantified. The methodology is widely applicable to different problem scenarios where several competing goals are considered simultaneously.

LA-UR 12-

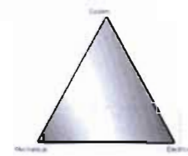
Strategic Future Data Collection for Complex Systems Based on Multiple Reliability Objectives

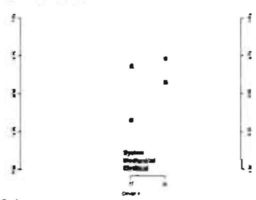


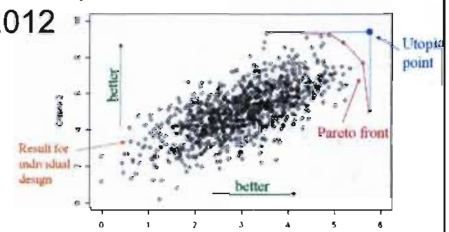
Christine Anderson-Cook
Los Alamos National Laboratory

Lu Lu & Jessica Chapman

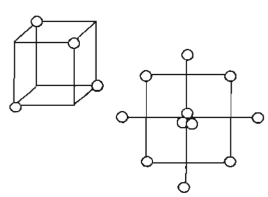
March 2012

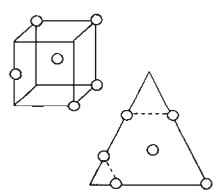


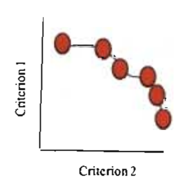




A Very Brief History of Design of Experiments







Textbook designs
- only some N
- regular regions
- good general performance

"Optimal" designs
- flexible N, region, criterion

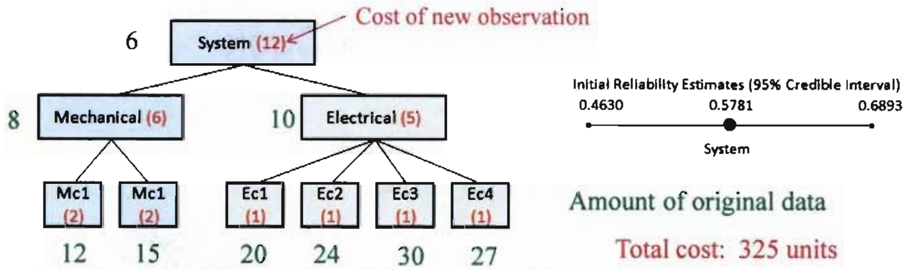
"Multiple Criteria" designs
- flexible N, region
- consider multiple objectives
- Pareto front based

Computer Power increasing dramatically

UNCLASSIFIED

Problem Statement

- Given the results of an existing reliability analysis based on multiple sources of data, what new data should we collect to **maximally improve** our estimation?

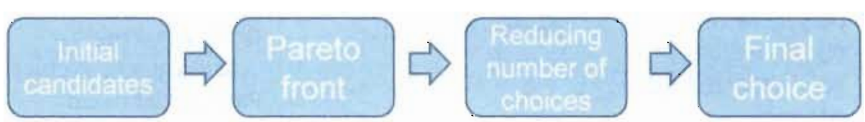


- What new data should we collect?
- What basis should we use for choosing?
- How do we justify what is best for our goals?

UNCLASSIFIED

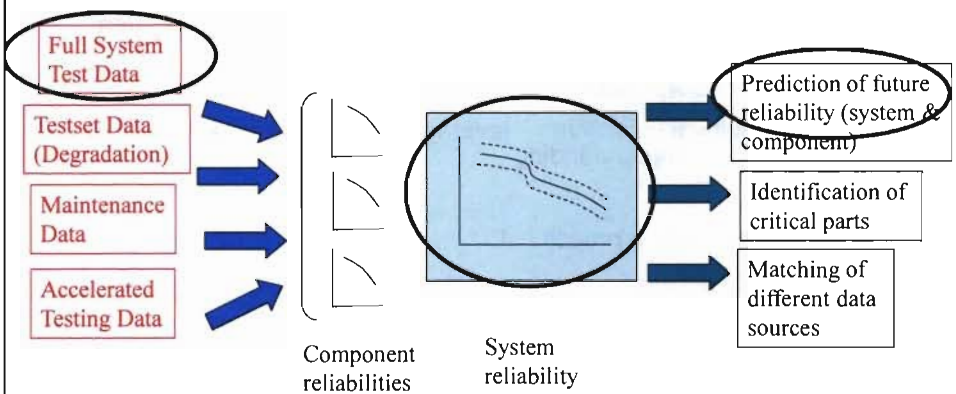
Outline

1. Introduce example – background on initial analysis + goals of new data collection
2. Overview of "Resource Allocation"
3. Overview of "Pareto Front Optimization"
4. Example revisited



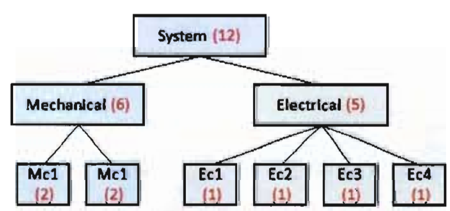
5. Conclusions

Bayesian Analysis with Multiple Data Types for Estimation of System Reliability



Basics of Multi-Level Reliability Model

- Simple model:
  - Assumes no aging of the system
  - All component reliabilities modeled as Binomial
  - System assumed to be Series (need all parts to work for the system to work, or system fails if at least one component fails)



P(Component i works) = p_i
P("Mechanical" works) = $p_M = p_{Mc1} \cdot p_{Mc2}$
P("Electrical" works) = $p_E = p_{Ec1} \cdot \ldots \cdot p_{Ec4}$
P("System" works) = $p_S = p_M \cdot p_E$

P(component fails) = $1 - p_i$
P("Mechanical" fails) = $1 - p_M$
P("System" fails) = $1 - p_S$

UNCLASSIFIED

Advantages of Data Combination Approach

- Uses data already available and thought to be relevant to predict reliability
- Improves precision of estimation with fewer destructive full-system tests
- Flexibility to incorporate partial information into model
- Component level reliabilities – leverage from different versions of system + better understanding

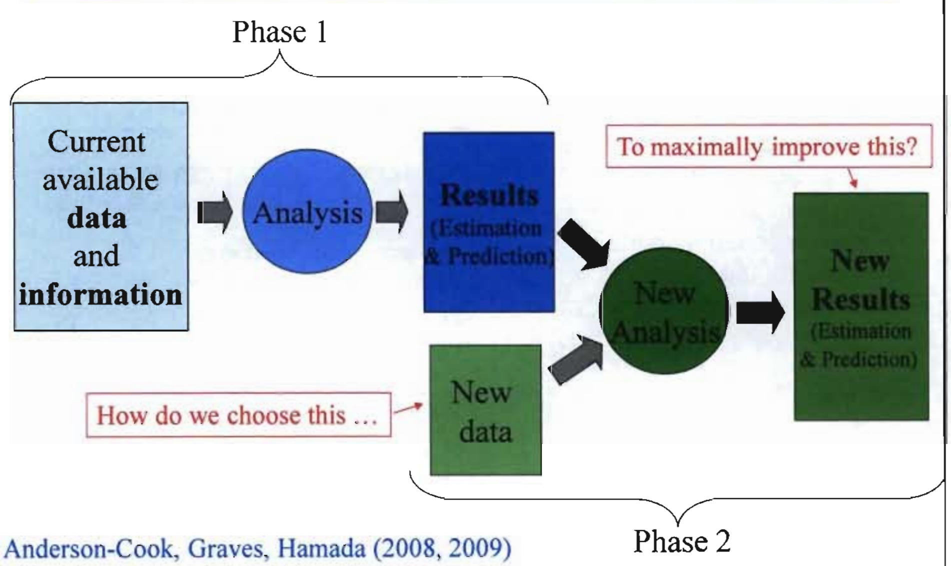
Why Use a Bayesian approach?

- Want to incorporate available expert opinion about reliability
- Easy combination of multiple levels of data with appropriate uncertainty propagation.

Disadvantage: More complex statistical method requiring more engineering knowledge to obtain results

UNCLASSIFIED

Resource Allocation – What are we trying to do?



Anderson-Cook, Graves, Hamada (2008, 2009)

UNCLASSIFIED

What does improvement mean here?

- Improve precision of estimates or prediction assuming model is correct (variance)
- Model discrepancy more precisely (estimate bias better)
- Look for problems / weaknesses / omissions in model (Unknown unknowns) (look for new bias)
- For multiple responses,
  need some prioritization / combination of responses to know how to quantify improvement

Budget for new data

Improve precision of existing model | Improve discrepancy estimate | Look for gaps in theory

Los Alamos NATIONAL LABORATORY EST. 1943

e 8

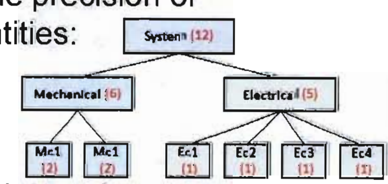
Operated by the Los Alamos National Security, LLC for the DOE/NNSA

NNSA

UNCLASSIFIED

Goal of New Data Collection for our Example

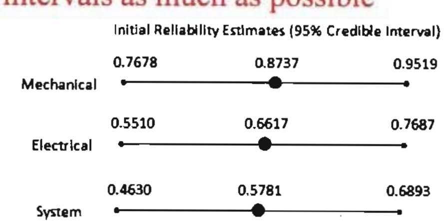
- Engineers would like to improve the precision of estimation for the following 3 quantities:
  - System reliability estimate
  - Mechanical Sub-system
  - Electrical Sub-system



- Focus on the width of the credible interval:

Goal: Reduce the width of each of these 3 intervals as much as possible

Baseline:



Width: 0.1841

Width: 0.2177

Width: 0.2263

Los Alamos NATIONAL LABORATORY EST. 1943

Slide 9

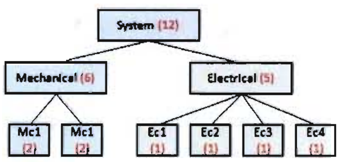
Operated by the Los Alamos National Security, LLC for the DOE/NNSA

NNSA

UNCLASSIFIED

Allocations Possible

25 possible allocations:
-All have total cost of 120 units
- Good variety of where data are collected



Results

Mc1	Mc2	Ec1	Ec2	Ec3	Ec4	Mechanical	Electronic	System	Alloc. #
0	0	0	0	0	0	0	0	10	1
0	0	0	0	0	0	20	0	0	2
0	0	0	0	0	0	0	24	0	3
0	0	0	0	0	0	10	0	5	4
0	0	0	0	0	0	0	12	5	5
0	0	0	0	0	0	10	12	0	6
15	15	0	0	0	0	0	0	5	7
0	0	15	15	15	15	0	0	5	8
8	7	0	0	0	0	5	0	5	9
0	0	9	8	6	7	5	0	5	10
0	0	9	8	6	7	0	6	5	11
8	7	0	0	0	0	0	6	5	12
8	7	9	8	6	7	0	0	5	13
16	14	0	0	0	0	10	0	0	14
8	7	9	8	6	7	10	0	0	15
0	0	18	16	12	14	0	12	0	16
8	7	9	8	6	7	0	12	0	17
8	7	0	0	0	0	5	12	0	18
0	0	9	8	6	7	5	12	0	19
0	0	9	8	6	7	10	6	0	20
8	7	0	0	0	0	10	6	0	21
30	30	0	0	0	0	0	0	0	22
0	0	30	30	30	30	0	0	0	23
15	15	15	15	15	15	0	0	0	24
17	13	18	16	12	14	0	0	0	25

UNCLASSIFIED

How to Evaluate New Data Before Collection

- After the initial analysis has been completed, we have all the posterior distributions for parameters

$$p_S,\ p_M,\ p_E,\ p_{Mc1},\ p_{Mc2},\ p_{Ec1},\ldots,p_{Ec4}$$

$$p(\theta|\mathbf{x}_1,\mathbf{x}_2) = \frac{p(\theta,\mathbf{x}_1,\mathbf{x}_2)}{p(\mathbf{x}_1,\mathbf{x}_2)} = \frac{p(\mathbf{x}_2|\theta)p(\theta|\mathbf{x}_1)}{p(\mathbf{x}_2|\mathbf{x}_1)}$$

- Approximate $p(\theta|\mathbf{x}_1)$ by treating θ as discrete and "making a histogram"
- For each (discrete) value of θ, approximate $p(\mathbf{x}_2|\theta)$ by averaging over the rows that correspond to that value of θ
- Approximate $p(\mathbf{x}_2|\mathbf{x}_1)$ by averaging over all rows

	θ	$P(\mathbf{x}_2 = 0\|\theta,\mathbf{x}_1)$	……	$P(x_2 = \mathbf{n}_2\|\theta,\mathbf{x}_1)$
$\dot{\theta}_1$	$\theta_{(1)}$	$P(\mathbf{x}_2 = 0\|\theta_{(1)},\mathbf{x}_1)$	……	$P(x_2 = \mathbf{n}_2\|\theta_{(1)},\mathbf{x}_1)$
	$\theta_{(2)}$	$P(\mathbf{x}_2 = 0\|\theta_{(2)},\mathbf{x}_1)$	……	$P(x_2 = \mathbf{n}_2\|\theta_{(2)},\mathbf{x}_1)$
$\dot{\theta}_2$	$\theta_{(3)}$	$P(\mathbf{x}_2 = 0\|\theta_{(3)},\mathbf{x}_1)$	……	$P(x_2 = \mathbf{n}_2\|\theta_{(3)},\mathbf{x}_1)$
	$\theta_{(4)}$	$P(\mathbf{x}_2 = 0\|\theta_{(4)},\mathbf{x}_1)$	……	$P(x_2 = \mathbf{n}_2\|\theta_{(4)},\mathbf{x}_1)$
	$\theta_{(5)}$	$P(\mathbf{x}_2 = 0\|\theta_{(5)},\mathbf{x}_1)$	……	$P(x_2 = \mathbf{n}_2\|\theta_{(5)},\mathbf{x}_1)$
$\vdots$	$\vdots$	$\vdots$	……	$\vdots$
$\dot{\theta}_N$	$\theta_{(M-3)}$	$P(\mathbf{x}_2 = 0\|\theta_{(M-3)},\mathbf{x}_1)$	……	$P(x_2 = \mathbf{n}_2\|\theta_{(M-3)},\mathbf{x}_1)$
	$\theta_{(M-2)}$	$P(\mathbf{x}_2 = 0\|\theta_{(M-2)},\mathbf{x}_1)$	……	$P(x_2 = \mathbf{n}_2\|\theta_{(M-2)},\mathbf{x}_1)$
	$\theta_{(M-1)}$	$P(\mathbf{x}_2 = 0\|\theta_{(M-1)},\mathbf{x}_1)$	……	$P(x_2 = \mathbf{n}_2\|\theta_{(M-1)},\mathbf{x}_1)$
	$\theta_{(M)}$	$P(\mathbf{x}_2 = 0\|\theta_{(M)},\mathbf{x}_1)$	……	$P(x_2 = \mathbf{n}_2\|\theta_{(M)},\mathbf{x}_1)$

Major advantage does not need **new MCMC** for different allocations

Slide 11

Details in Chapman, Morris and Anderson-Cook (2010)

UNCLASSIFIED

Process for Selecting a Best Design

- The process for finding a best design for our specific goals can be summarized by a multi-stage algorithm:

1. Evaluate all the collections of allowable new data allocations, and measure the three criteria for each.
2. Construct the Pareto front, which consists of all allocations which are not inferior to (*Pareto dominated* by) any others
3. Select a best allocation from the Pareto front which best suits the needs of the experimenter.

Los Alamos NATIONAL LABORATORY EST. 1943
Operated by the Los Alamos National Security, LLC for the DOE/NNSA

UNCLASSIFIED

Slide 12

NNSA

UNCLASSIFIED

Initial:
0.1841 0.2177 0.2263

Results of Analysis for 25 Allocations

Mc1	Mc2	Ec1	Ec2	Ec3	Ec4	Mechanical	Electronic	System	Alloc. #	Mechanical	Electrical	System
0	0	0	0	0	0	0	0	10	1	0.182896	0.208562	0.212798
0	0	0	0	0	0	20	0	0	2	0.155939	0.217593	0.216135
0	0	0	0	0	0	0	24	0	3	0.186184	0.188425	0.205978
0	0	0	0	0	0	10	0	5	4	0.167084	0.212652	0.213953
0	0	0	0	0	0	0	12	5	5	0.183102	0.196996	0.208728
0	0	0	0	0	0	10	12	0	6	0.169521	0.201595	0.208982
15	15	0	0	0	0	0	0	5	7	0.160029	0.212583	0.211654
0	0	15	15	15	15	0	0	5	8	0.184335	0.193014	0.20547
8	7	0	0	0	0	5	0	5	9	0.163004	0.212495	0.212505
0	0	9	8	6	7	5	0	5	10	0.176321	0.202387	0.209188
0	0	9	8	6	7	0	6	5	11	0.184579	0.196233	0.207769
8	7	0	0	0	0	0	6	5	12	0.171135	0.204599	0.20955
8	7	9	8	6	7	0	0	5	13	0.172838	0.202965	0.208023
16	14	0	0	0	0	10	0	0	14	0.150264	0.217789	0.214795
8	7	9	8	6	7	10	0	0	15	0.158434	0.206712	0.20864
0	0	18	16	12	14	0	12	0	16	0.186341	0.185205	0.202329
8	7	9	8	6	7	0	12	0	17	0.17274	0.192642	0.203601
8	7	0	0	0	0	5	12	0	18	0.166135	0.201575	0.207661
0	0	9	8	6	7	5	12	0	19	0.178806	0.19274	0.205274
0	0	9	8	6	7	10	6	0	20	0.169961	0.199462	0.207632
8	7	0	0	0	0	10	6	0	21	0.159256	0.209364	0.211409
30	30	0	0	0	0	0	0	0	22	$\boxed{0.143072}$	0.217661	0.213209
0	0	30	30	30	30	0	0	0	23	0.186722	$\boxed{0.18047}$	$\boxed{0.19775}$
15	15	15	15	15	15	0	0	0	24	0.163415	0.197407	0.203219
17	13	18	16	12	14	0	0	0	25	0.161928	0.196507	0.20274

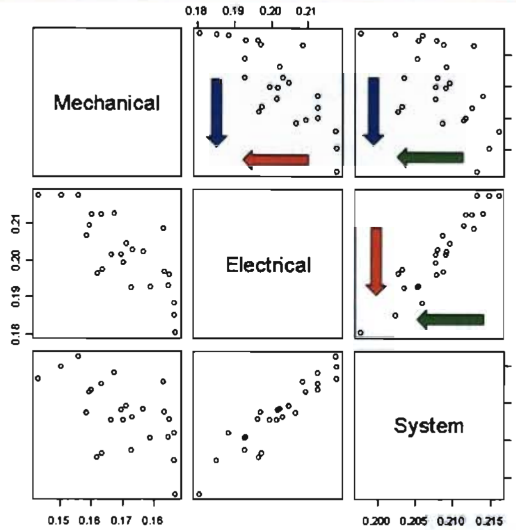
Slide 13

UNCLASSIFIED

Credible Interval Results for 25 allocations

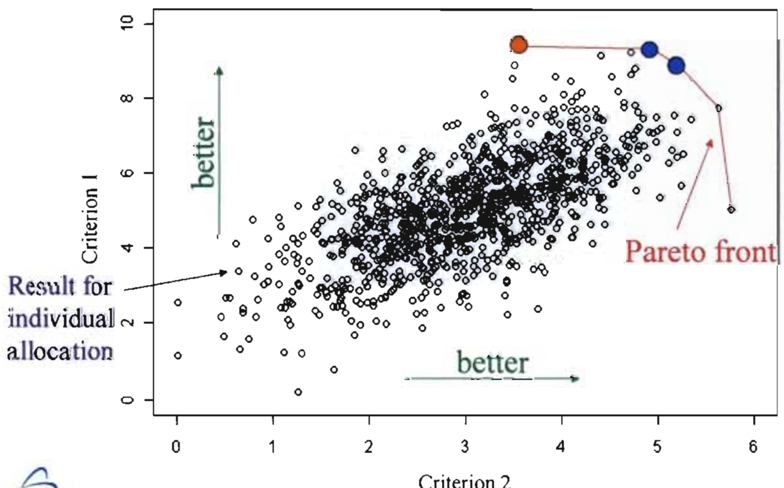
Recall: Goal is to minimize the width of the credible interval for each of:

- System
- Mechanical
- Electrical



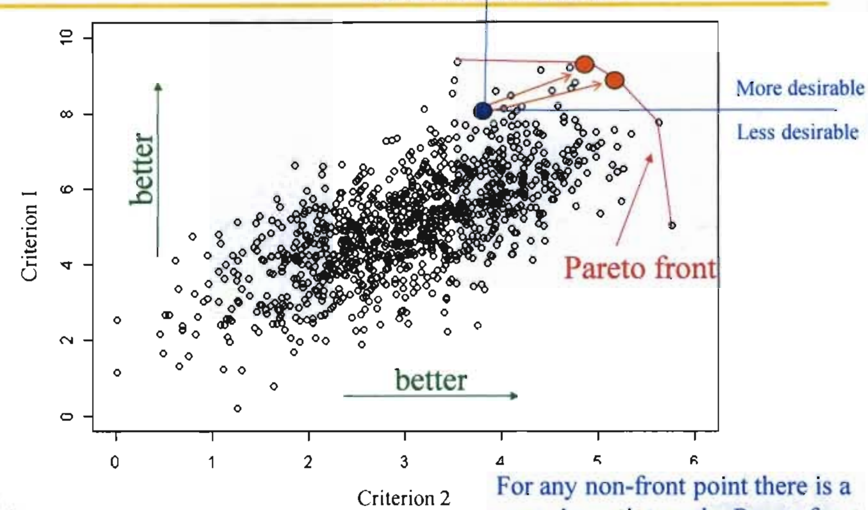
UNCLASSIFIED

The Weakness of Single Criterion Optimization



UNCLASSIFIED

The Pareto Front Approach to Eliminate Non-Contenders



For any non-front point there is a **superior** point on the Pareto front

Los Alamos NATIONAL LABORATORY

UNCLASSIFIED

Operated by the Los Alamos National Security, LLC for the DOE/NNSA

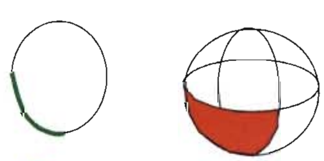
NNSA

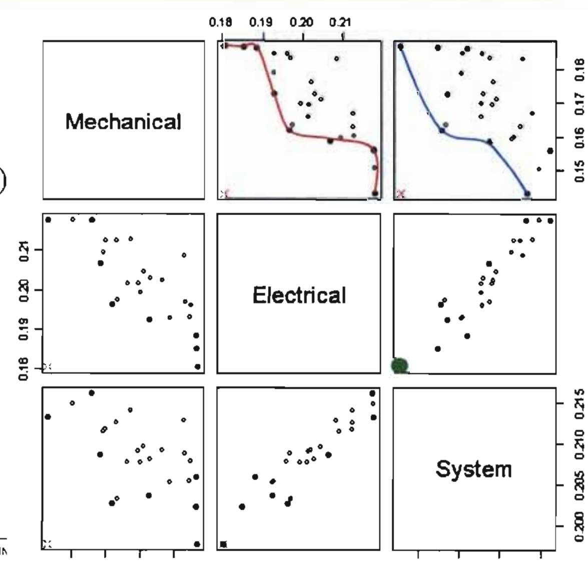
UNCLASSIFIED

Pareto Front for 25 allocations

Divides possible allocations into two groups:

- on Pareto Front (9 pts)
- not on front (16 pts)



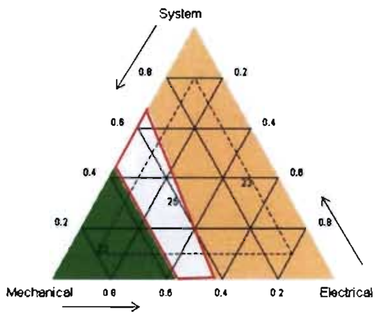


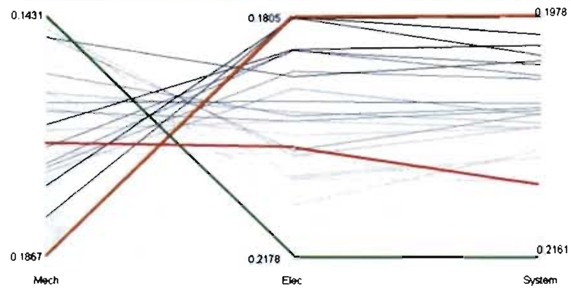
Los Alamos NATIONAL LABORATORY

Operated by the Los Alamos National Security, LLC for the DOE/NN

Finding Best Allocation for All Weightings of Criteria (Additive)

$$Desirability = w_S CW_S + w_M CW_M + w_E CW_E$$





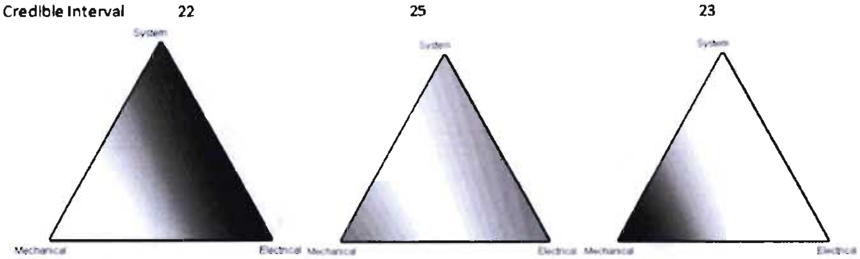
Mc1	Mc2	Ec1	Ec2	Ec3	Ec4	Mechanical	Electronic	System	Alloc. #
30	30	0	0	0	0	0	0	0	22
0	0	30	30	30	30	0	0	0	23
15	15	15	15	15	15	0	0	0	24
17	13	18	16	12	14	0	0	0	25

Lu, Anderson-Cook and Robinson, 2011

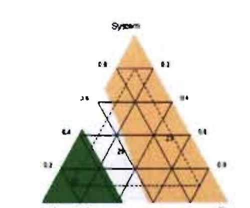
Efficiency Plot – Comparing to the Best Possible

Additive Combination of Criteria

Credible Interval 22 25 23



○100% ○75% ●50% ●25% ●0%



Lu & Anderson-Cook, 2012

Final Choice of Allocation

Based on Additive Combination of 3 Criteria:
-Allocations 22, 23, 25 are promising

If **system reliability** really is most important, then select 23

For **overall robustness** compared to best at any weight – choose 25

Mc1	Mc2	Ec1	Ec2	Ec3	Ec4	Mechanical	Electronic	System	Alloc. #	Mechanical	Electrical	System
8	7	9	8	6	7	0	12	0	17	0.17274	0.192642	0.203601
30	30	0	0	0	0	0	0	0	22	0.143072	0.217661	0.213209
0	0	30	30	30	30	0	0	0	23	0.186722	0.18047	0.19775
17	13	18	16	12	14	0	0	0	25	0.161928	0.196507	0.20274
									Initial:	0.1841	0.2177	0.2263
									% reduction:	12.0%	9.7%	10.4%
									Max % reduction individually :	22.3%	17.1%	12.6%

Conclusions

- Looking at multiple characteristics is often a more realistic way of selecting a best choices (life is rarely as simple is just wanting to do well on one characteristic)
- Different allocations have different advantages and risks – select criteria to consider which best capture the important considerations for your decision.
- The Pareto front approach can divide possible solutions into (1) those consider further and (2) those to eliminate, because they are dominated by other better choices
- Once the Pareto front has been selected, there are multiple ways of selecting the final solution – but the key is to examine and understand the trade-offs between the choices
- The Pareto front approach is suitable for all types of multiple response optimizations – we just need to specify what criteria we wish to optimize over. Search algorithms exist for how to find a best solution (not just find the best from among a list (Lu, Anderson-Cook and Robinson, 2011)

Other Applications of Pareto Front Approach for Design (in various publications)

- Example 1 (Screening Experiment):
  - D-optimality [maximize |X'X|]
  - Good estimation of pure error [maximize df_{PE}]
  - Good estimation of lack of fit [maximize tr(R'R) /(m-p)]
- Example 2 (Robust Parameter Design Experiment):
  - Good estimation of terms affecting the mean [max D_s-mean]
  - Good estimation of terms affecting the variance [max D_s-variance]
  - Size of experiment [min N]
- Example 3 (Split Plot Design):
  - Good estimation of terms when WP to SP variance ratio is unknown [max D(1), max D(10)]
  - Size of experiment [min N]
  - Number of Whole Plots [min #WP]

References

- Anderson-Cook, C.M., Chapman, J., Lu, L. (2011) "Selecting a Best Allocation of New Data for Improving Estimation Precision of System and Sub-System Reliability using Pareto Fronts: A Demonstration" Los Alamos National Laboratory Technical Report, LAUR 11-05479
- Anderson-Cook, C.M., Graves, T.G., Hamada, M.S. (2009) "Resource Allocation for Reliability of a Complex System with Aging Components" **Quality and Reliability Engineering International** 25 481-494.
- Chapman, J., Morris, M., Anderson-Cook, C.M. (2011) "A Computationally Efficient Strategy for Evaluating the Estimation Improvement for Candidates in a Resource Allocation Study" Los Alamos National Laboratory Technical Report LA-UR 10-04501
- Lu L., 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** (to appear)
- Lu, L., Anderson-Cook, C.M., Robinson, T.J. (2011) "Optimization of Designed Experiments Based on Multiple Criteria Utilizing a Pareto Frontier" **Technometrics** 53 353-365.