SAND NO. 2013-8393C

SAND2013-8393C

Sensitivity Analyses for Multi-Objective Combinatorial Optimization (MOCO) Problems

Alex Dessanti

Sandia National Laboratories

2013 INFORMS Annual Meeting
October 6, 2013
Minneapolis, MN

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Contributors

- Sponsor: Mr. Shatiel Edwards

  Program Executive Office Ground Combat Systems (PEO GCS)

- Multi-Organization Team: PEO GCS, Sandia National Laboratories, Booz Allen Hamilton, Armament Research Development and Engineering Center

Sandia National Laboratories

Briefing Overview

The purpose of this briefing is to provide an overview of one proposed methodology to perform a multi-objective combinatorial optimization sensitivity analysis and discuss other potential techniques.

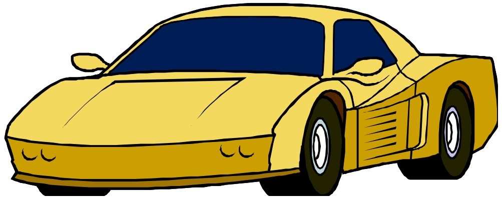
- **Motivation**
- **Problem**
- **Background**
  - **Multi-Objective Optimization**
  - **Combinatorial Optimization**
  - **Sensitivity Analyses**
- **Approaches Considered**
- **Stepwise Regression**
  - **Notional Cost Example**
  - **Notional Cost and Weight Example**
  - **Findings**
- **Future Work**
- **Questions/Discussion**

Motivation

- Multiple reasons to be interested in sensitivity analyses for multi-objective combinatorial optimization problems
  - Sensitivity analyses provide guidance regarding which parameters are most influential to the output of a problem
  - Complex problems that are difficult to fully understand at first glance (such as MOCO), can use sensitivity analyses to help provide insights for better understanding of results
  - Pressure to reduce costs and increase efficiencies necessitates better identification of areas that drive a problem's solution
- Literature is lacking with respect to sensitivity analyses for multi-objective combinatorial optimization
  - Multi-objective optimization inherently represents the sensitivity of the outputs to one another
  - However, it does not necessarily indicate "Why were these solutions chosen?" or "What parameters have the greatest impact on the solutions?"

Problem

- Whole System Trades Analysis Tool (WSTAT) is a multi-objective combinatorial optimization tool
- WSTAT can be applied to explore design decisions for essentially any system or system of systems
- Problem examined here explores the *notional* trade space for a car
  - Design decisions include choice of engine, transmission, body style, etc.
- Trades are made with respect to five dimensions of stakeholder value (each dimension can have constituent measures that are combined through a weighted sum)
  - Performance
  - Purchase Cost
  - Lifecycle Costs
  - Market Share
  - Safety



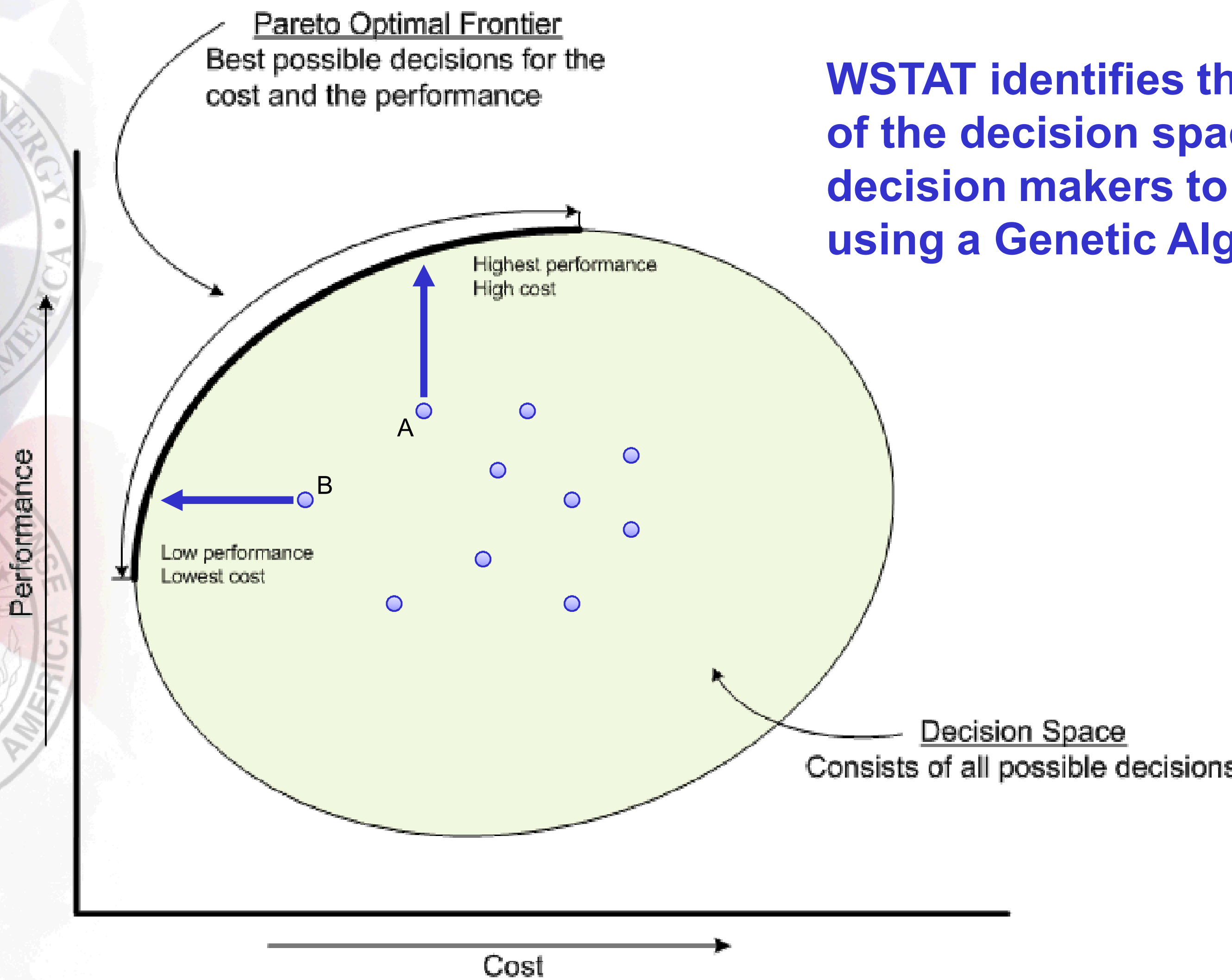
Sandia National Laboratories

Background

- Multi-objective optimization is used to determine the best choices for acceptable ranges of budget, benefit, risk, etc.
- Combinatorial optimization is used to determine the best collection of objects (technologies, activities, etc.) from groups of possible alternatives
  - Designing vehicles
  - Planning courses of action
  - Planning investments
- Sandia has combined and progressed these disciplines to provide decision makers with an advanced, defensible analytical capability

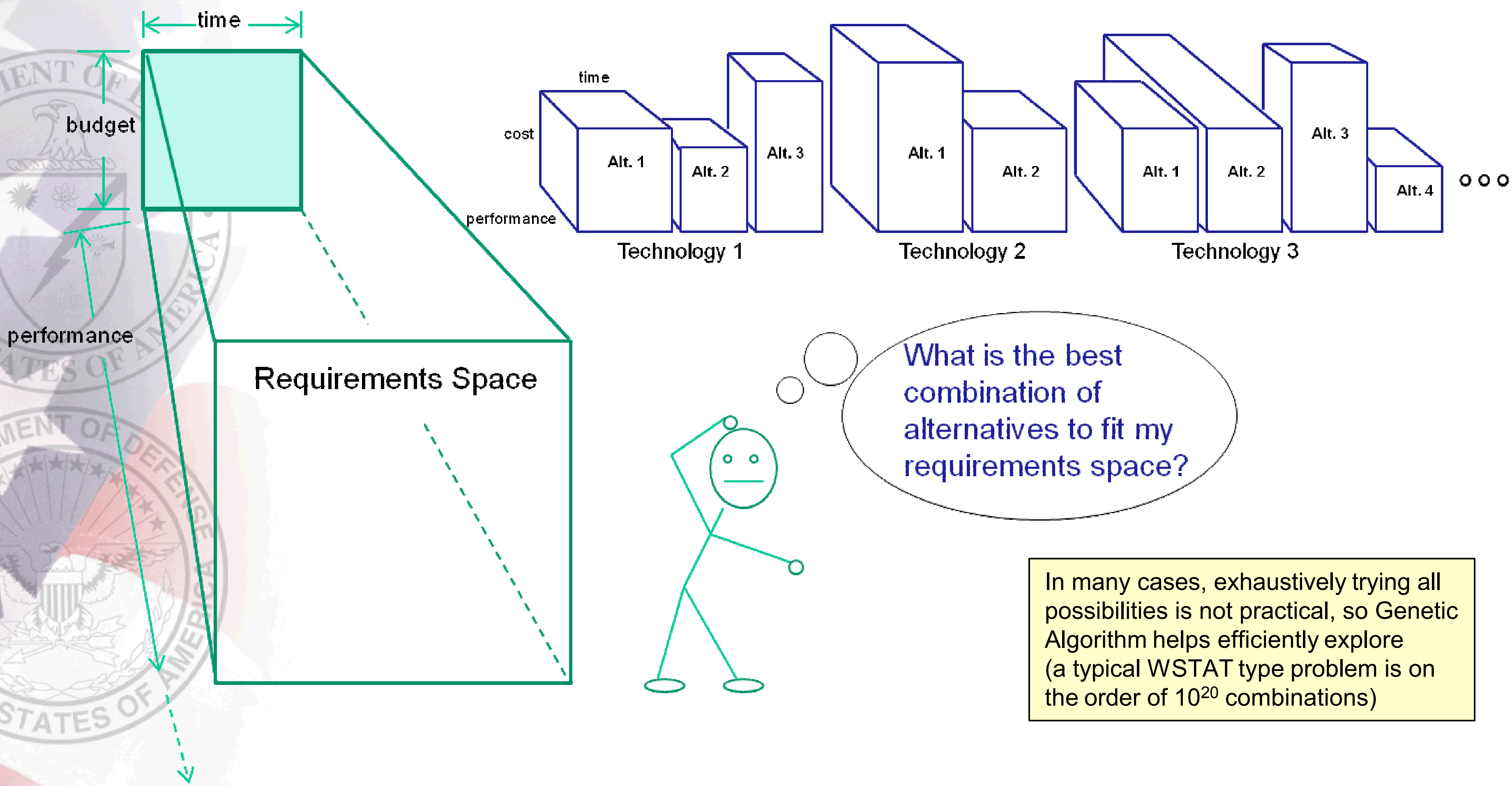
WSTAT

Multi-Objective Optimization



WSTAT identifies the best part of the decision space for decision makers to consider using a Genetic Algorithm

Combinatorial Optimization



In many cases, exhaustively trying all possibilities is not practical, so Genetic Algorithm helps efficiently explore (a typical WSTAT type problem is on the order of 10^{20} combinations)

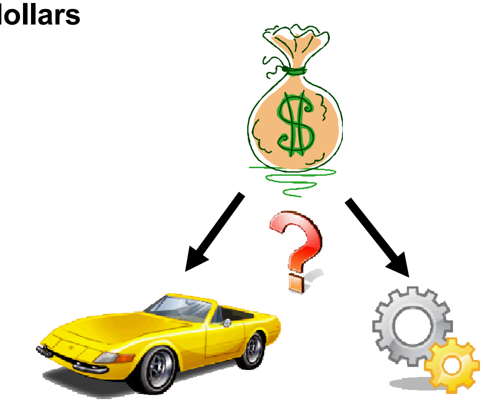
For us, Combinatorial Optimization addresses the problem of “which ones?”

Sensitivity Analyses

- **Attempt to identify which model inputs are most correlated with the variability in the model outputs**
  - **Assume a stronger correlation indicates greater importance of that input (driving output)**
  - **Improving on that parameter should lead to better results quicker than some other, less correlated, parameter**
    - **For technology cost inputs and overall system cost as an output, lowering the cost of a highly correlated technology should have a greater impact on the overall system cost**
- **Specifically, here we are trying to answer the question of "Which technology areas should a decision maker concentrate on?"**
  - **For example, is the body style choice or the transmission choice more important to the resulting cost of the system?**
  - **Want to identify which technology selections and attributes the optimal trade space is most sensitive to, so decision makers know where to focus attention/investment dollars**
- **Considerations for sensitivity analysis**
  - **Ease of use**
  - **Usefulness**
  - **Accuracy**
  - **Consistent with a multi-objective combinatorial optimization**

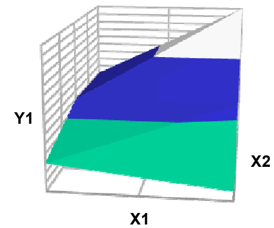
Transmission Options
4-Speed Manual
5-Speed Manual
6-Speed Manual
6-Speed Automatic

Body Options
Coupe
Sedan
Wagon

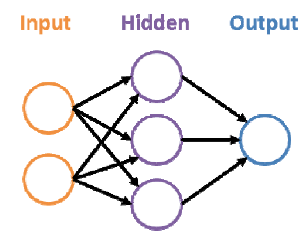


Approaches Considered

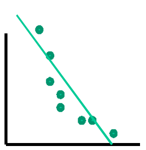
- Response Surface Methodology
  - n^{th} degree polynomial model
  - Good for modeling relationships between an output and several inputs
  - By examining two-dimensional slices of the n-dimensional surface, one can explore how an output changes with respect to a single input while all other inputs are held constant



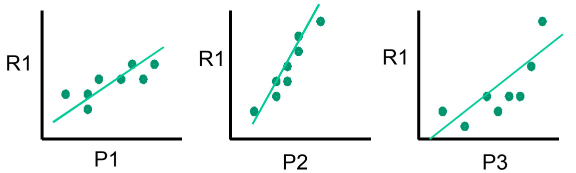
- Artificial Neural Networks
  - Non-linear model
  - Good at finding patterns in data when inputs and outputs have a complex relationship



- Correlation Coefficients
  - Can find correlation coefficients between all the inputs and outputs
  - Provides insight into the direction and strength of the relationship
  - Does not predict the magnitude of the impact an input has on the output



- Stepwise Regression
  - Linear model
  - Utilized for first sensitivity analysis attempt



Stepwise Regression

- Method for building a linear model from input parameters that approximates the output results
  - Adds or removes variables that improve the model most and repeats until no further improvement occurs
  - Magnitude of the coefficient of determination (R^2) indicates the degree of sensitivity of the results to the parameter
- It is a purely mechanistic method that does not include any information about the original model – it has been criticized for being “data dredging”
  - Can lead to spurious results
  - Often difficult to extend model beyond dataset used to create it due to overfitting
- Here extended to multi-objective optimization
  - Given a Pareto frontier…
    - Inputs are the values selected for each solution in the optimized set (e.g. the costs of engines A, B, and C)
    - Outputs are the various rolled-up values for a solution (e.g. the total cost)

Stepwise regression is a straightforward, unbiased (albeit naïve) method for determining linear correlations between model inputs and outputs as well as the relative magnitude of the impact

Sandia National Laboratories

Stepwise Regression Example – Cost

Inputs (28)
1937 Samples

Body Cost ($)	Engine Cost ($)	Transmission Cost ($)	Tire Cost ($)	Seat Cost ($)	Radio Cost ($)	Speaker Cost ($)	• • •
13846	3212	2887	397	955	122	962	
14200	3595	4487	371	872	216	1052	
10140	3150	2469	480	957	121	1069	
12624	2831	4536	391	787	179	919	
14062	4679	2102	235	630	134	801	
11895	4416	4974	415	769	293	1171	
11707	2579	3355	369	997	160	1015	
11779	3527	4919	274	577	150	850	
13265	4370	2195	429	696	258	883	
14588	4373	3058	421	925	147	810	
12052	3690	4479	640	958	214	796	
12643	3310	4719	365	885	138	1057	
10236	2794	4232	471	626	183	1003	
11237	3286	2177	467	635	299	988	
13811	4367	4403	326	558	210	934	
13770	3024	2525	551	920	278	1041	
10095	4036	2946	394	538	209	835	
10563	3382	3675	339	992	229	1125	
13735	4455	3984	346	893	142	911	

⋮

Output

Purchase Cost ($)
20,282
43,270
29,450
21,097
30,137
49,741
28,889
40,910
48,691
37,203
41,424
33,458
46,245
23,687
36,838
28,458
48,924
35,182
39,487

Input and Output data is imported to Minitab which is used to perform Stepwise Regression

Question – the cost of which technology selection has the greatest impact on the system's Purchase Cost?

Sandia National Laboratories

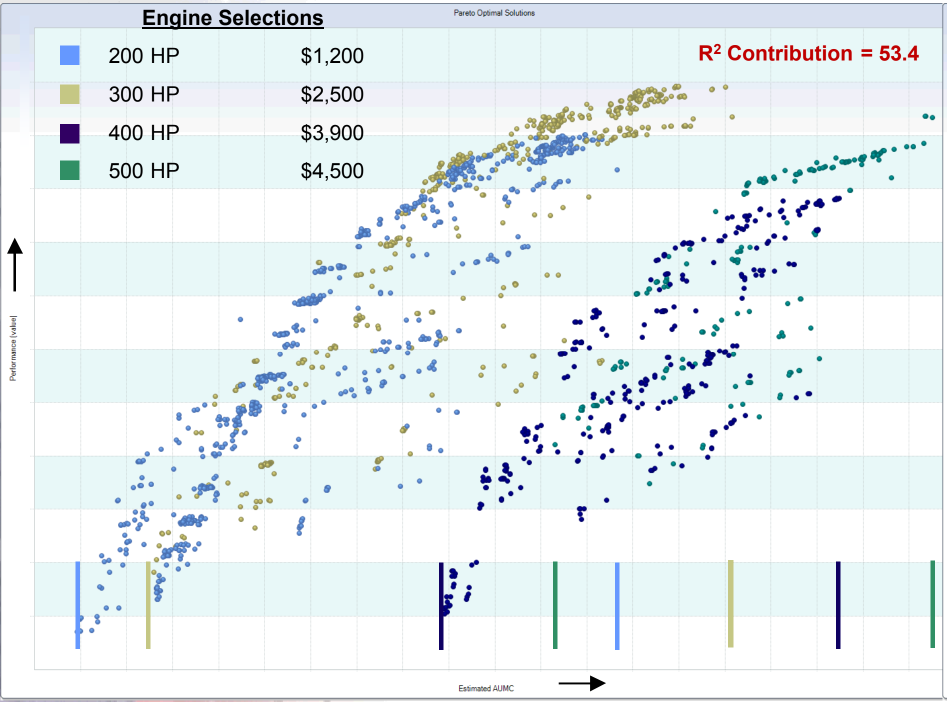
Stepwise Regression Example – Cost

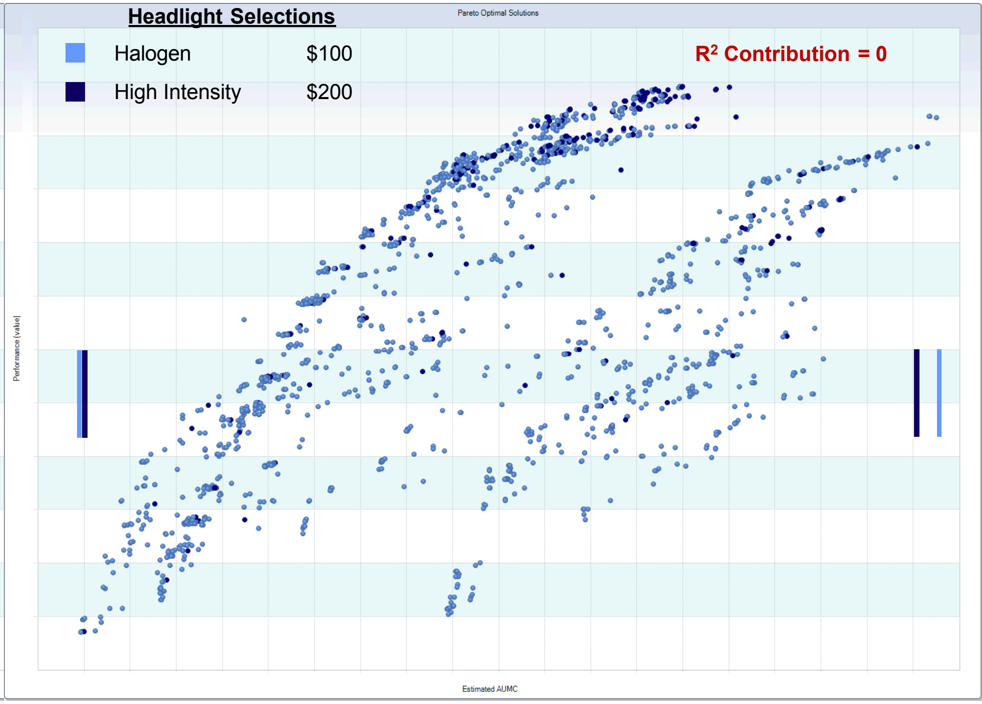
Technology	R^2	ΔR^2
Engine	53.40	53.40
Body Style	83.79	30.39
Transmission	92.34	8.55
Heating/Cooling	94.47	2.13
Tires	96.20	1.73
Brakes	97.74	1.54
Speakers	98.68	0.94
Seats	99.14	0.46
Radios	99.45	0.31
Other	99.79	0.34

- Note that most of the variance in the system Purchase Cost is captured by this linear model (final R^2 of 99.79)
- Costs of 9 technologies have the greatest impact on the system cost
  - Engine and Body Style selections have the largest impacts
  - Might indicate to the decision maker that in order to reduce system cost, the cost of these technologies should be monitored most closely

Sandia National Laboratories

Stepwise Regression Example – Cost





- **Driving Technology Selections (High ΔR^2)**
  - **Across range of observed system costs, one can see that cheaper options are first seen at a lower system cost than more expensive options (see Engine Above)**
  - **Significant spread in option costs**
- **Non-Driving Technology Selections (Low ΔR^2)**
  - **All options picked across range of system costs (see Headlight above)**
  - **Minimal spread in option costs**

Sandia National Laboratories

Stepwise Regression Example – Cost and Weight

Inputs (56)

Technology	Parameter
Body Style	Cost
Engine	
Transmission	
Tires	
Brakes	
Emergency Brake	
Seats	
Heating/Cooling	
Speakers	
Radios	
Headlights	

⋮

Technology	Parameter
Body Style	Weight
Engine	
Transmission	
Tires	
Brakes	
Emergency Brake	
Seats	
Heating/Cooling	
Speakers	
Radios	
Headlights	

⋮

Output

Purchase Cost

Question – does the weight of any of the technologies impact the system Purchase Cost?

Stepwise Regression Example – Cost and Weight

Cost Inputs Only

Technology	R^2	ΔR^2	Attribute
Engine	53.40	53.40	Cost
Body Style	83.79	30.39	Cost
Transmission	92.34	8.55	Cost
Heating/Cooling	94.47	2.13	Cost
Tires	96.20	1.73	Cost
Brakes	97.74	1.54	Cost
Speakers	98.68	0.94	Cost
Seats	99.14	0.46	Cost
Radios	99.45	0.31	Cost
Other	99.79	0.34	-

Cost and Weight Inputs

Technology	R^2	ΔR^2	Attribute
Engine	53.40	53.40	Cost
Body Style	83.79	30.39	Cost
Transmission	92.34	8.55	Weight
Heating/Cooling	94.97	2.63	Weight
Tires	96.49	1.52	Cost
Brakes	97.87	1.38	Cost
Speakers	98.78	0.91	Cost
Seats	99.18	0.40	Cost
Radios	99.45	0.27	Cost
Other	99.84	0.39	-

- Top 9 contributors identified in both cases are the same (both order and approximate contribution)
  - Top 3 contributors have exact same ΔR^2
- For two technology areas (Transmission and Heating/Cooling), the model identified the weight attribute rather than the cost attribute as a primary contributor to the Purchase Cost output (with approximately the same impact as original example)
- While it is possible that the weight could have an impact on cost, there are other explanations for why this might be occurring
  - Cost and Weight inputs happen to be correlated for those technologies
  - Fewer different technology options selected, so limited data from which to infer relationships

Stepwise Regression Findings

- Characteristics of technologies that are correlated to a Pareto optimal output
  - More than one technology option selected
  - Selected options have a wide range of values
  - Selected options have larger magnitude values
  - Technology selections appear in certain ranges of the output
- Methodology
  - Simple to perform using statistical software
  - Handles multi-objective combinatorial optimization problems
  - Can help identify driving inputs
  - Upon creating the model, results need to be further examined
    - Ensure each technology/attribute identified as having a large impact can be explained
    - Technologies can appear to have a large contribution when in fact that contribution is an artifact of the data used to create the regression model

Future Work

- **Improve understanding of Stepwise Regression results**
- **Explore other techniques for identifying driving inputs**
  - Response Surface Methodology
  - Artificial Neural Networks
  - Correlation Coefficients
- **Investigate sensitivity to Priority Weightings**
  - Different stakeholders have different priorities
  - Important aspect of this optimization problem
    - 5 Dimensions are not weighted, but constituent measures are (especially for Performance)
    - Want to be able to show how sensitive results are to priority weightings to help mediate discussions between different stakeholders
  - Want to set up Design of Experiments
  - Issues will require some thought
    - Run time/computational demands
    - Identifying changes from one model run to another

Questions/Discussion