Interactive evaluative model trading for resilient systems decisions

By

Adam M. Ross, Matthew E. Fitzgerald, and Donna H. Rhodes

14th Annual Conference on Systems Engineering Research (CSER)

March 23, 2016

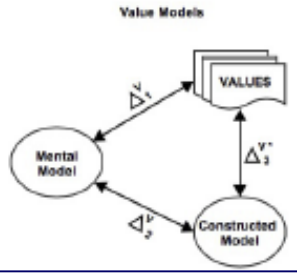
University of Alabama-Huntsville

Huntsville, AL

Model Choice and Tradeoffs

- Models increasingly used to predict system performance, and costs and benefits
- Every model is an abstraction from reality; important to understand implications of embedded assumptions
- Sensitivity analyses essential, yet many studies resource constrained and neglect
- Since assumptions impact results of models, choices of model parameters important for "within" model sensitivity, AND most importantly, choices of models themselves
- Correctness of most appropriate model may be *uncertain*
- Resilient decisions are those that perform well in spite of such uncertainties and sensitivities (e.g. "correct" and appropriate model choice)

Constructed Model v. Mental Model		Mental Model v. Values: Low Δ_1 (truthful)	Mental Model v. Values: High Δ_1
	Low Δ_2 (trust)	**Correct Trust:** Trust present. Correct representation of values.	**Type I error:** Trust present. Incorrect representation of values.
	High Δ_2	**Type II error:** Trust not present. Correct representation of values.	**Justified Mistrust:** Trust not present. Incorrect representation of values.



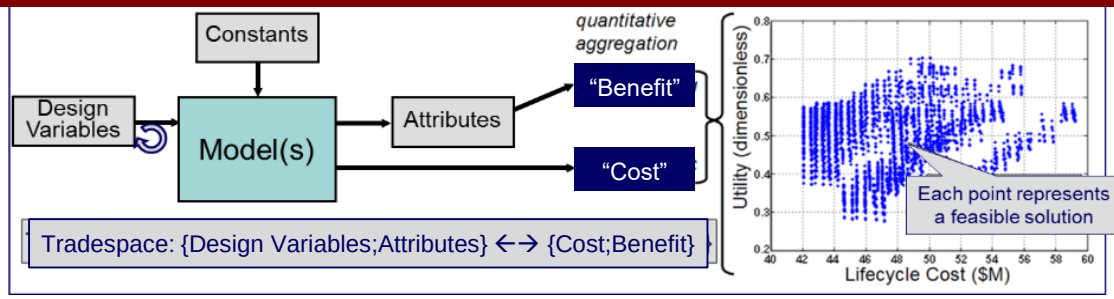
Preliminary research was done to trade "within model" sensitivities in value models, investigating the potential for interaction in refining value model parameter choices*

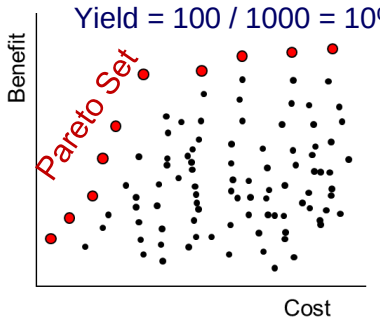
*Ricci, N., Schaffner, M.A., Ross, A.M., Rhodes, D.H., and Fitzgerald, M.E., "Exploring Stakeholder Value Models Via Interactive Visualization," 12th CSER, Redondo Beach, CA, March 2014.

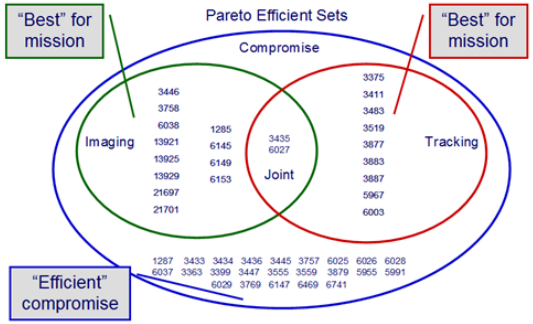
How can we perform effective "sensitivity" analyses in terms of model choice so as to enable resilient decision outcomes?

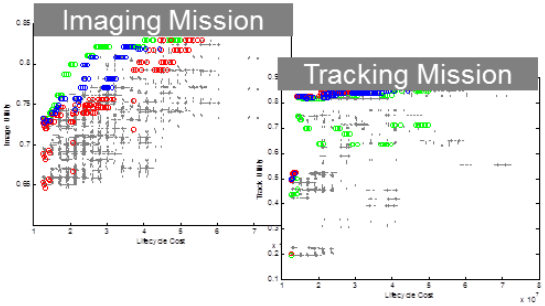
Tradespace Concepts

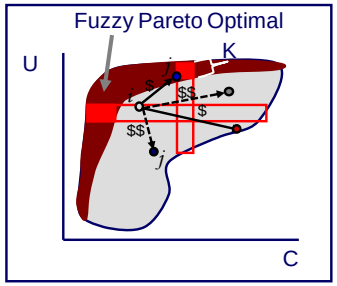
Used tradespace exploration to see consequences on potential solutions of model choices
Used IVTea Suite software (internal MIT) for visuals and analyses







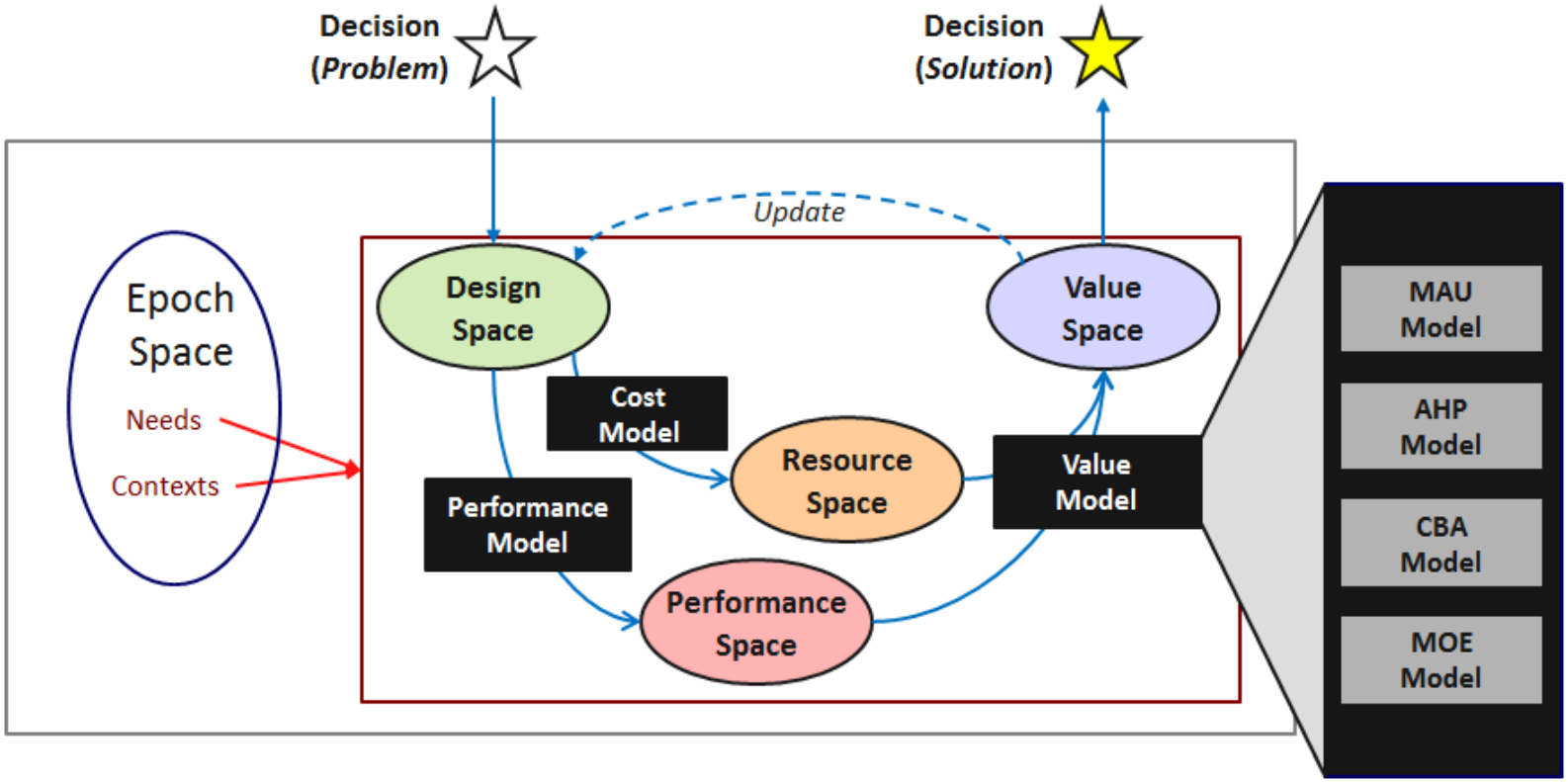




- **Scatterplot**: 2D plot where x,y position of point indicates scores in those dimensions (often benefit and cost)
- **Yield**: Fraction of design space considered valid
- **Pareto Set**: Non-dominated (most efficient) set of designs for given objectives (i.e. improvement in one objective results in degradation of other objective)
- **Joint/Compromise**: Joint designs are in multiple individual Pareto sets; compromise designs are in combined objective Pareto set but not individual Pareto sets
- **K% Fuzzy Pareto Set**: Designs that are included in Pareto set when allowing for within K% of Pareto frontier

Previously: Value Model Trading

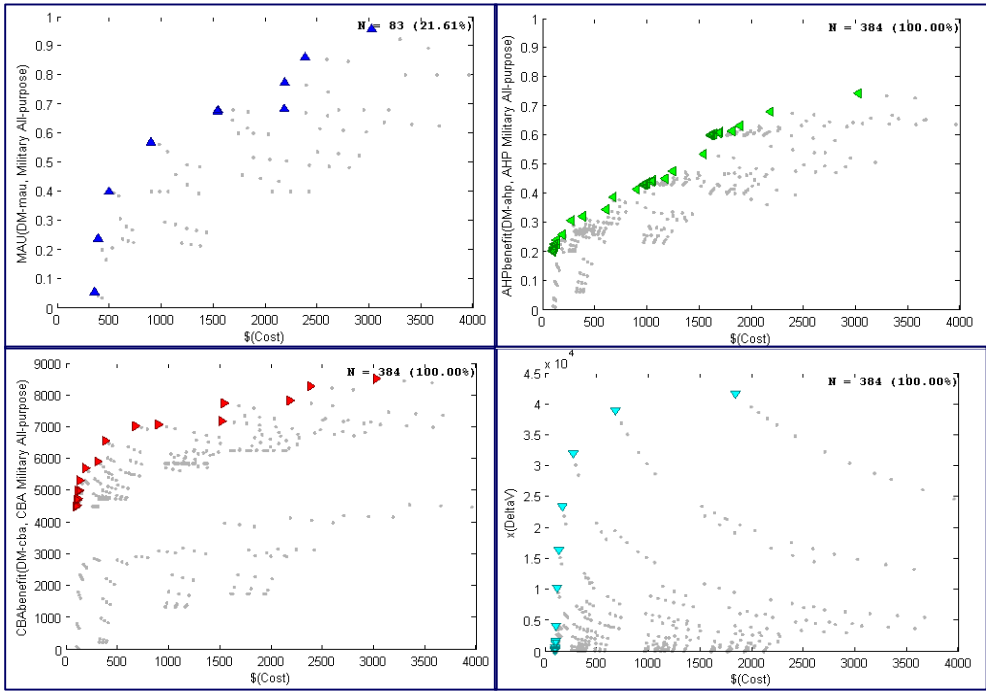
Design-Value Loop: Value model is just as important as performance and cost



Since choosing / tuning a value model is hard – can we build confidence in decisions or trust in models by examining differences between them?

Value Model Trade Outcomes

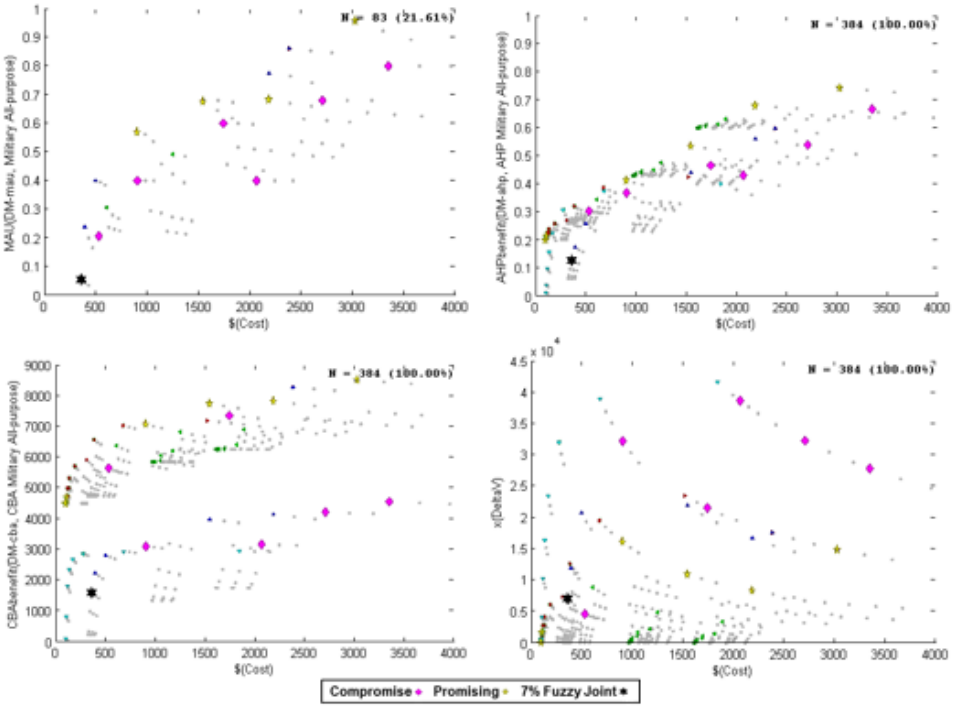
- 4 value models
  - Multi-Attribute Utility
  - Analytic Hierarchy Process
  - Cost-Benefit Analysis
  - Measure of Effectiveness



Ross, A.M., Rhodes, D.H., and Fitzgerald, M.E., "Interactive Value Model Trading for Resilient Systems Decisions," 13th Conference on Systems Engineering Research, Hoboken, NJ, March 2015.

Value Model Trade Outcomes

- 4 value models
  - Multi-Attribute Utility
  - Analytic Hierarchy Process
  - Cost-Benefit Analysis
  - Measure of Effectiveness
- Demonstrated
  - Impact of model choice on "good" alternatives
  - Cross-model comparison to build trust by identifying joint, "promising" (nearly joint), and compromise efficient alternatives

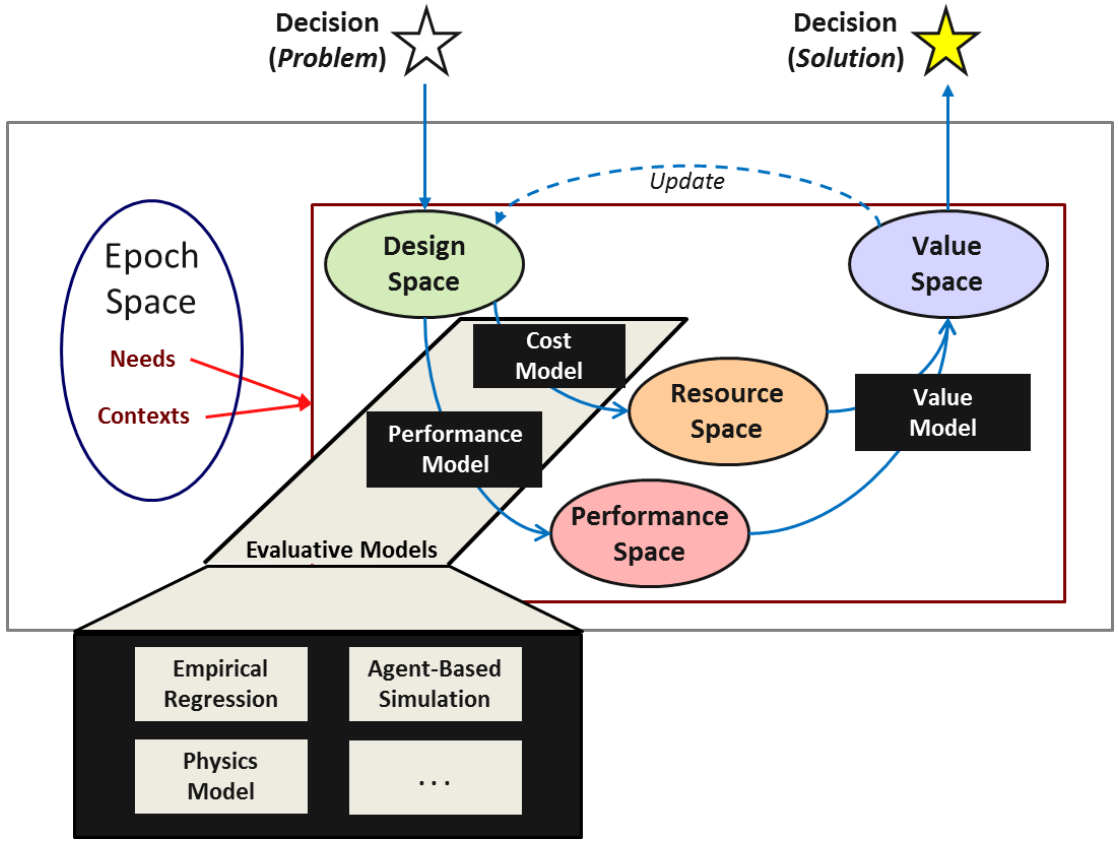


Ross, A.M., Rhodes, D.H., and Fitzgerald, M.E., "Interactive Value Model Trading for Resilient Systems Decisions," 13th Conference on Systems Engineering Research, Hoboken, NJ, March 2015.

Evaluative Model Trading

What about the other models?

- Evaluative models are also chosen and impact the decision
- Traditionally: tradeoff between fidelity and time
  - "Best" model may be subjective
  - For conceptual design, there may be no validated model



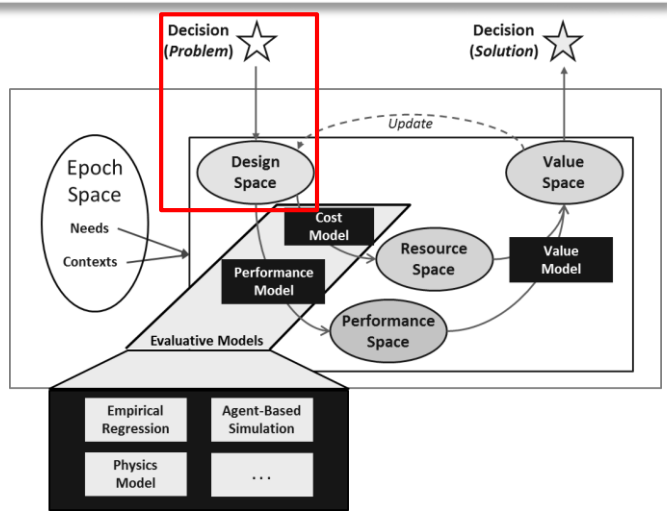
Can we also leverage multiple evaluative models in order to develop new insights into the system?

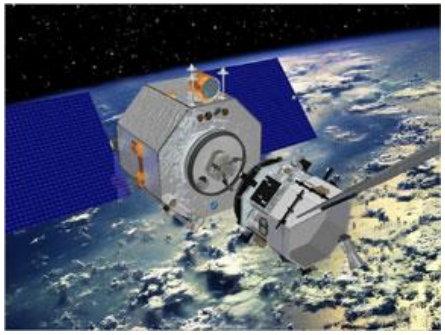
Reintroducing Space Tug

- Same case as prior study

General purpose vehicle to intercept, interact with, and accelerate other vehicles

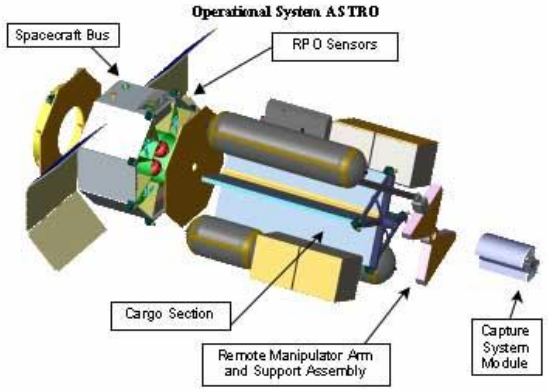
Decision problem: *what would be a valuable design for such a vehicle?*





Attributes
- ❑ Delta-V
- ❑ Capability
- ❑ Response time

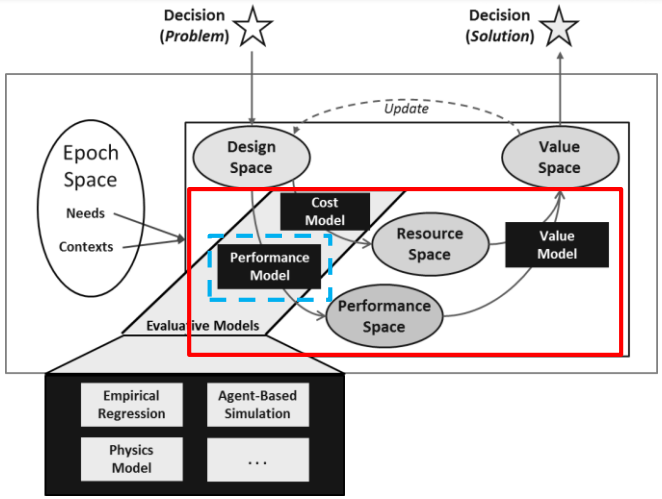
Design Space (N=384)*
- ❑ Manipulator Mass
  - Low (300kg)
  - Medium (1000kg)
  - High (3000 kg)
  - Extreme (5000 kg)
- ❑ Propulsion Type
  - Storable bi-prop
  - Cryogenic bi-prop
  - Electric (NSTAR)
  - Nuclear Thermal
- ❑ Fuel Load - 8 levels
- ❑ Design for Change (DfC)
  - None, level-1, level-2



*Fitzgerald, M.E., Ross, A.M., and Rhodes, D.H., "Assessing Uncertain Benefits: a Valuation Approach for Strategic Changeability (VASC)," INCOSE International Symposium 2012, Rome, Italy, July 2012.

Space Tug Models

- Simple (low-fidelity) formulation
- Value model fixed to MAU



Performance Model (base*)

Propulsion System Mass	$M_p = m_{p0} + m_{pf} M_f$
Bus Mass	$M_b = M_p + m_{bf} M_c$
Dry Mass	$M_d = M_b + M_c$
Wet Mass	$M_w = M_d + M_f$

Delta-V	$\Delta v = g\, I_{sp}\, \ln(M_w / M_d)$
Capability	*Table lookup*
Speed	*Table lookup*

Propulsion System	I_{sp} (sec)	Base Mass m_{p0} (kg)	Mass Fract. m_{pf}	High Impulse
Storable biprop	300	0	0.12	Y
Cryo	450	0	0.13	Y
Electric	3000	25	0.30	N
Nuclear	1500	1000	0.20	Y

*McManus, H., and Schuman, T., "Understanding the Orbital Transfer Vehicle Tradespace," AIAA-2003-6370, Sept. 2003

Value Model (MAU)

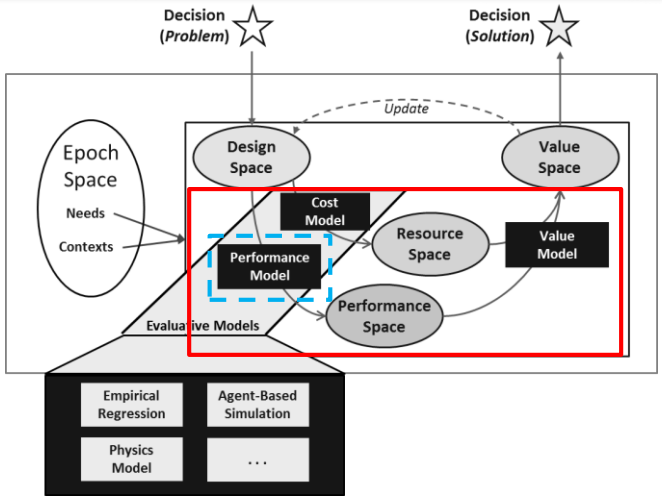
$$U(\hat{X}) = \frac{[\prod_{i=1}^{n}(K \cdot k_i \cdot U_i(X_i) + 1)] - 1}{K} \quad \text{where} \quad K = -1 + \prod_{i=1}^{n}(K \cdot k_i + 1)$$

Cost Model

$$C = c_w\, M_w + c_d\, M_d$$

Space Tug Performance Models

- Simple (low-fidelity) formulation
- Expanded to 4 implementations



#1 Performance Model (base*)

Propulsion System Mass	$M_p = m_{p0} + m_{pf} M_f$
Bus Mass	$M_b = M_p + m_{bf} M_c$
Dry Mass	$M_d = M_b + M_c$
Wet Mass	$M_w = M_d + M_f$
Delta-V	$\Delta v = g\, I_{sp}\, \ln(M_w / M_d)$
Capability	*Table lookup*
Speed	*Table lookup*

Propulsion System	I_{sp} (sec)	Base Mass m_{p0} (kg)	Mass Fract. m_{pf}	High Impulse
Storable biprop	300	0	0.12	Y
Cryo	450	0	0.13	Y
Electric	3000	25	0.30	N
Nuclear	1500	1000	0.20	Y

#2 Performance Model (new speed)

"accel"

#3 Performance Model (new material)

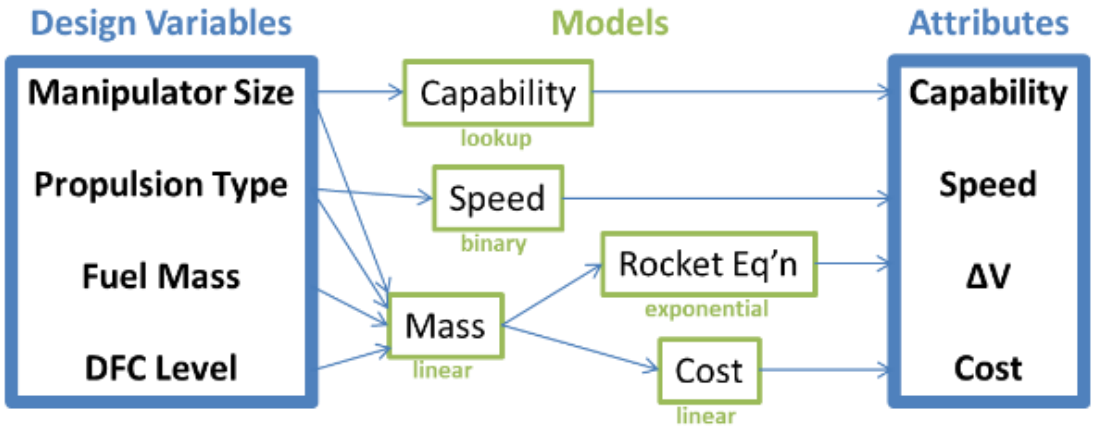
"carbon"

#4 Performance Model (new speed and material)

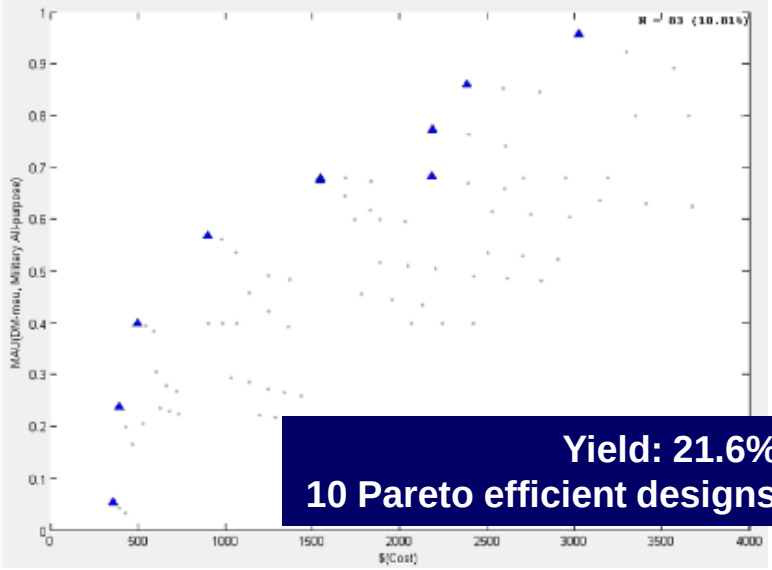
"both"

Implementation #1

- Uses “base” model as in prior study
  - Capability: lookup table
  - Speed: binary lookup
  - Delta-V: rocket equation
  - Cost: linear



Tradespace matches MAU tradespace in Value Model Trading study



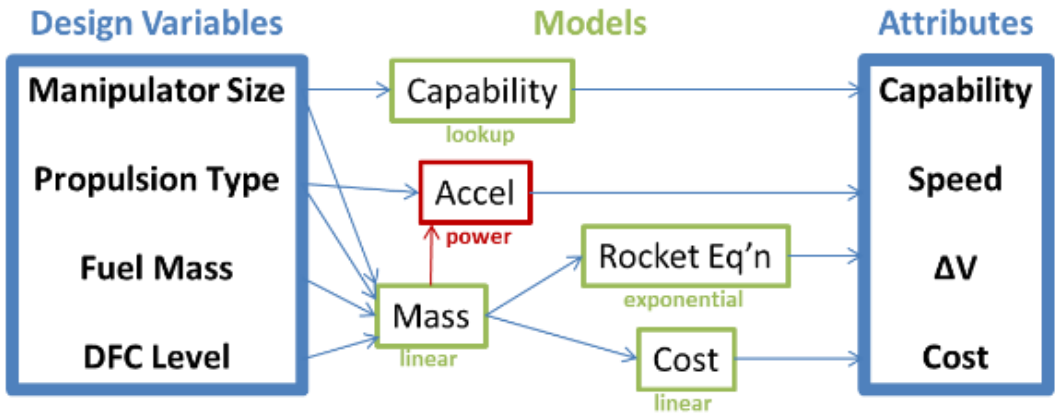
Yield: 21.6%
10 Pareto efficient designs

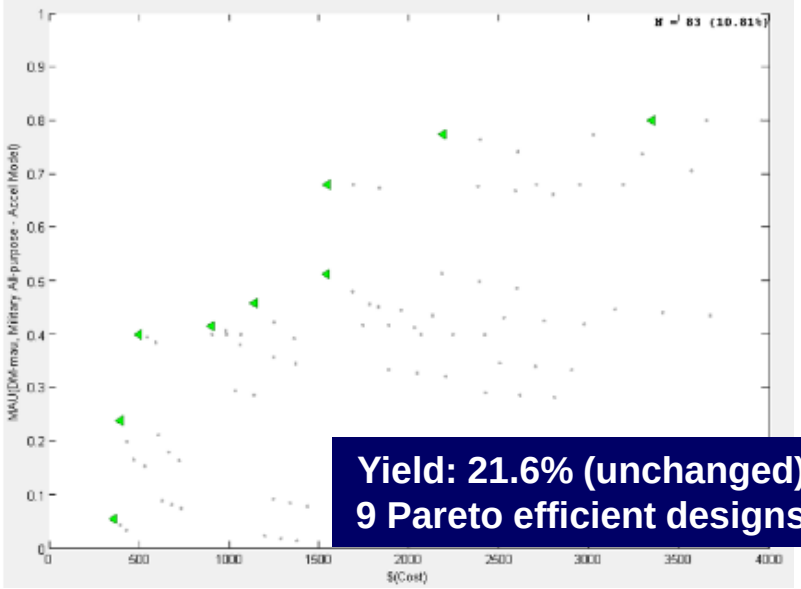
Implementation #2

- Speed model changed
  - Speed now evaluated as integrated acceleration over time, not binary
  - Classic fidelity increase

$$a = \frac{F}{M_w}$$

Propulsion Type	Thrust (N)
Bipropellant	500
Cryogenic	10,000
Electric	0.5
Nuclear	10,000

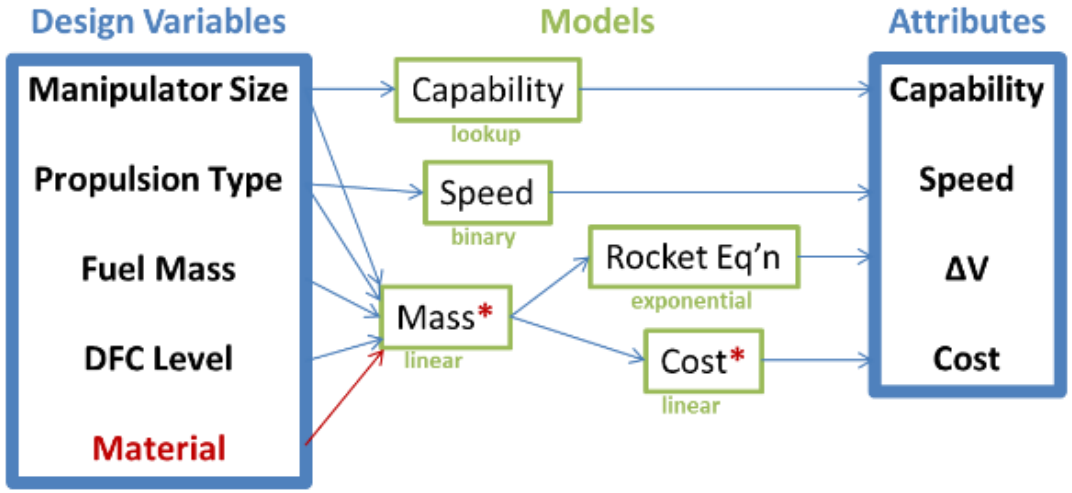


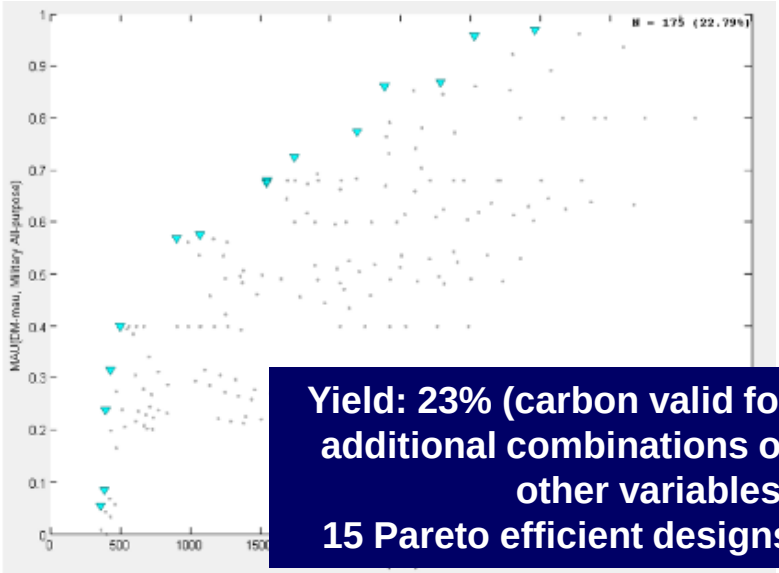


Implementation #3

- New design choices allowed
  - Material added as a design variable
  - "Representational" fidelity increase
  - Doubles size of the tradespace

	Bus mass (fraction of payload) m_{bf}	Dry mass cost rate c_d ($K/kg)
Aluminum (base)	1	150
Carbon (new)	0.6	200

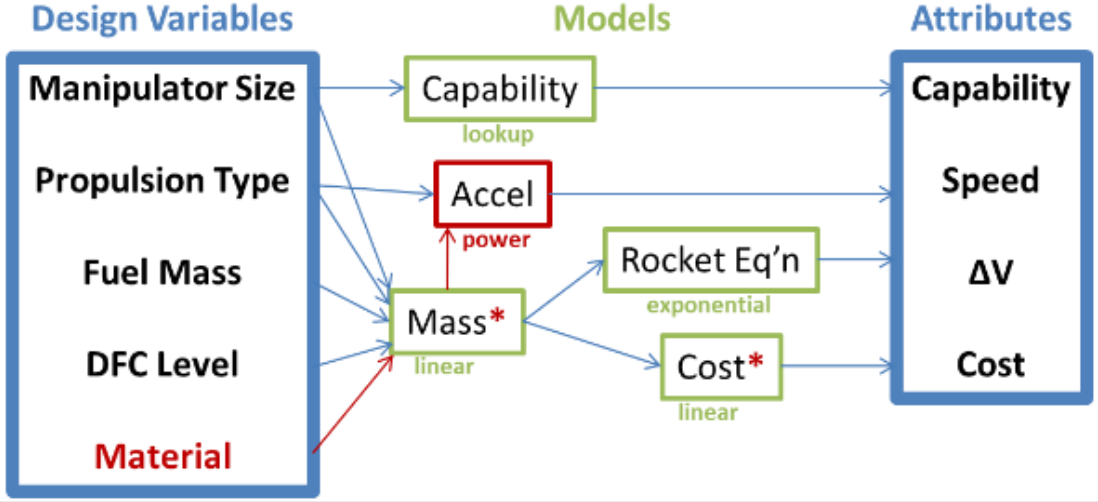


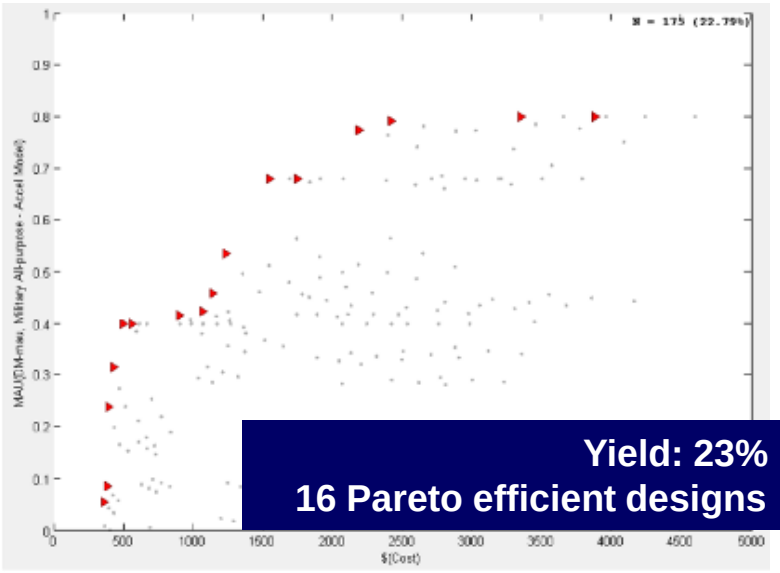


Yield: 23% (carbon valid for additional combinations of other variables)
15 Pareto efficient designs

Implementation #4

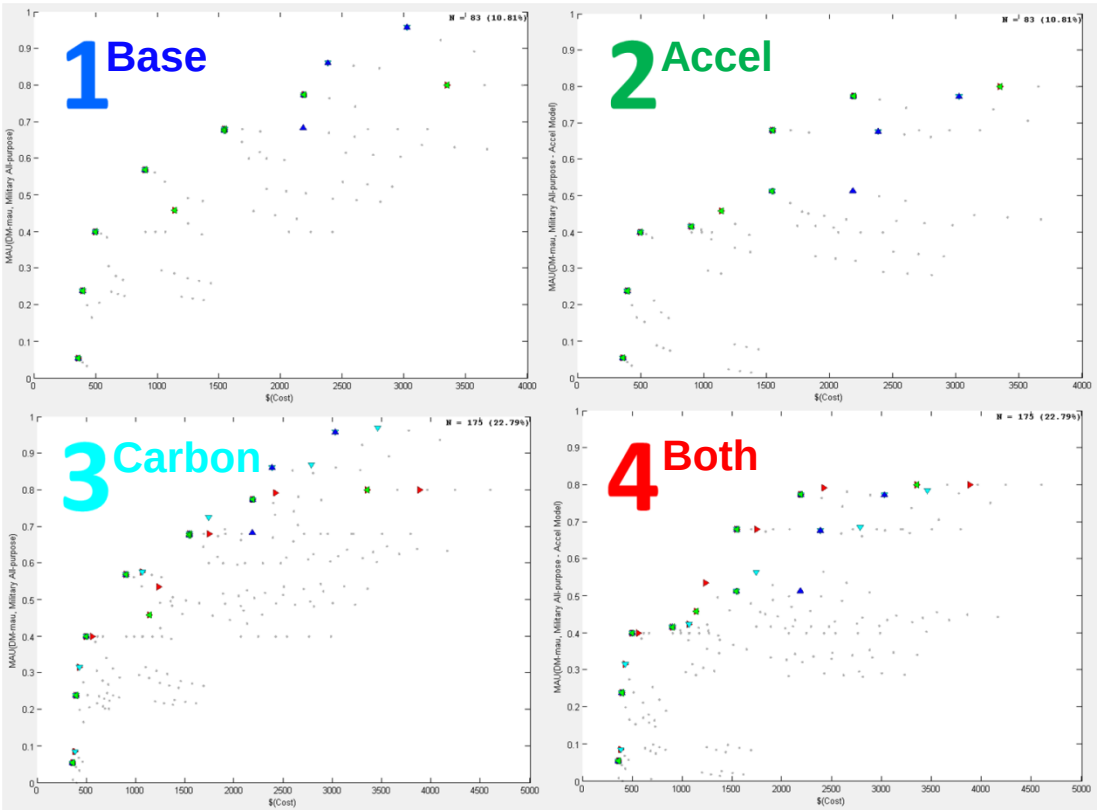
- Both speed and material changes included
  - Models mix-and-match modeling choices

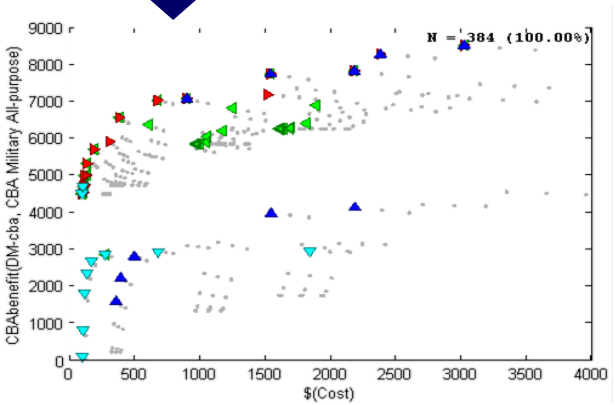




4 Implementations, 4 Pareto Sets

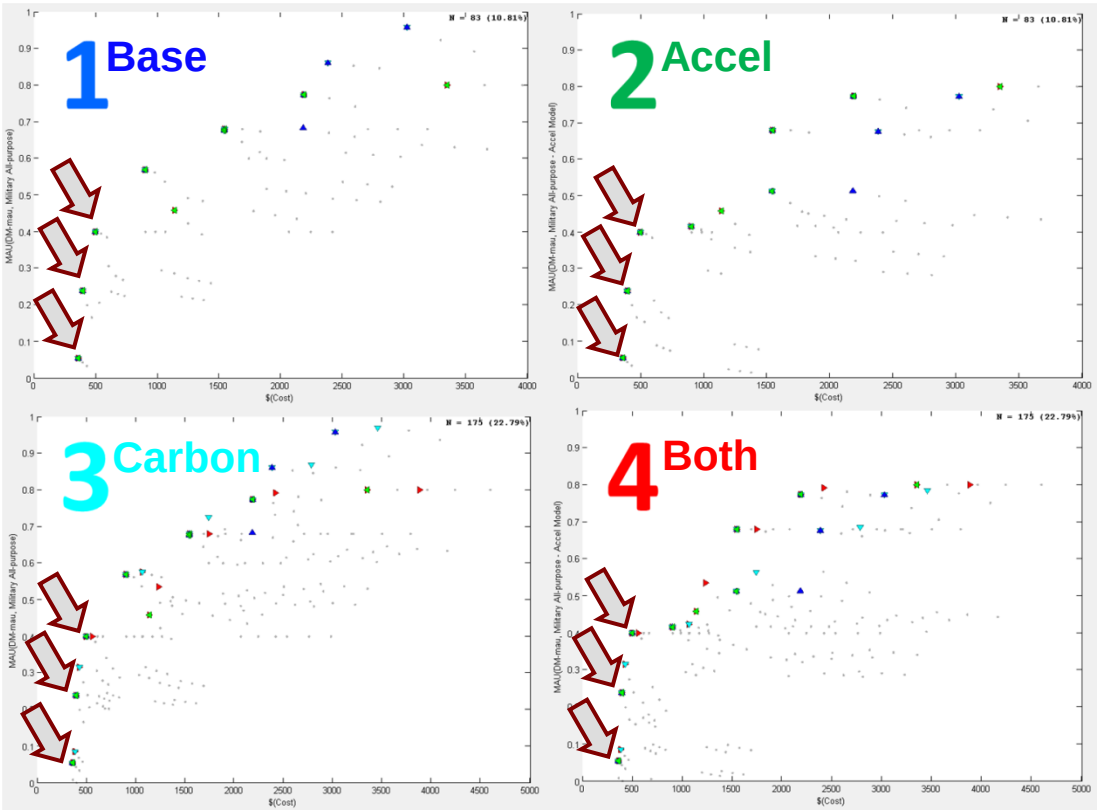
- Much more overlap than for value trading
  - 'Fit for purpose' is more aligned than different subjective value interpretations

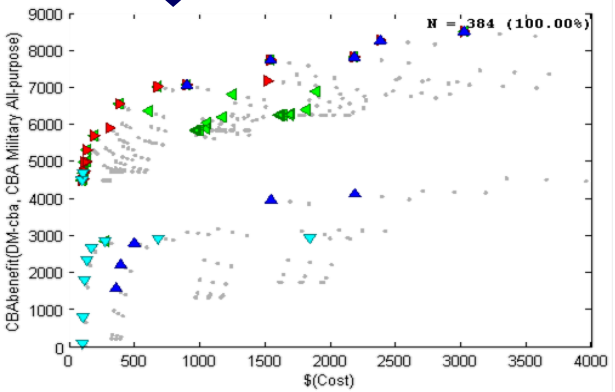




4 Implementations, 4 Pareto Sets

- Much more overlap than for value trading
  - 'Fit for purpose' is more aligned than different subjective value interpretations





At least three designs in the low-cost region are the same across all four models

- Adding "fuzziness" increases the size of the sets but shows the same patterns, focus here will be on the actual front

Cross-Model Comparison

- Six categories of efficient designs across the four models

	Base	Accel	Carbon	Both		Carbon	Both
	Design ID (Aluminum)	Model 1	Model 2	Model 3	Model 4	Design ID (Carbon)	Model 3	Model 4
A	52	√	√	√	√	436	√	√
	53	√	√	√	√	437	√	√
	63	√	√	√	√	447	√	√
B	54	√	√	√	√	438		√
	87	√	√	√	√	471		√
	119	√	√	√	√	503		√
C	86		√		√	470		√
	120		√		√	504		√
D	96	√		√		480	√	
	128	√		√		512	√	
E	127	√				511		
F	95	√	√	√		479	√	

Aluminum Designs **Carbon Designs**

√ **On Pareto front in that model implementation**

What insights about the Space Tug system can we learn from examining the designs in each category?

Most Efficient Designs

- Cat. **A** is always efficient, in all models with both materials
  - These are the low-cost, robust designs we saw before
  - Medium payloads, electric or nuclear propulsion

	Base	Accel	Carbon	Both		Carbon	Both
	Design ID (Aluminum)	Model 1	Model 2	Model 3	Model 4	Design ID (Carbon)	Model 3	Model 4
A	52	√	√	√	√	436	√	√
	53	√	√	√	√	437	√	√
	63	√	√	√	√	447	√	√
B	54	√	√	√	√	438		√
	87	√	√	√	√	471		√
	119	√	√	√	√	503		√

- Cat. **B** is efficient except with carbon material under the binary speed model (3), as lower weight is not rewarded:
  - Delta-V near or above the "max value" point of the utility function, not improved by lower mass
  - Binary speed model does not tie speed to mass

Cost increase of carbon needs to be offset by AT LEAST one of these

Sometimes Efficient Designs

- Cat. **C** is efficient only under acceleration model
  - These are Cat. **B** designs with slightly different amounts of fuel
  - Different "sweetspot" in the amount of fuel mass when mass is penalized by the speed model

		Base	Accel	Carbon	Both		Carbon	Both
	Design ID (Aluminum)	Model 1	Model 2	Model 3	Model 4	Design ID (Carbon)	Model 3	Model 4
C	86		√		√	470		√
	120		√		√	504		√
D	96	√		√		480	√	
	128	√		√		512	√	

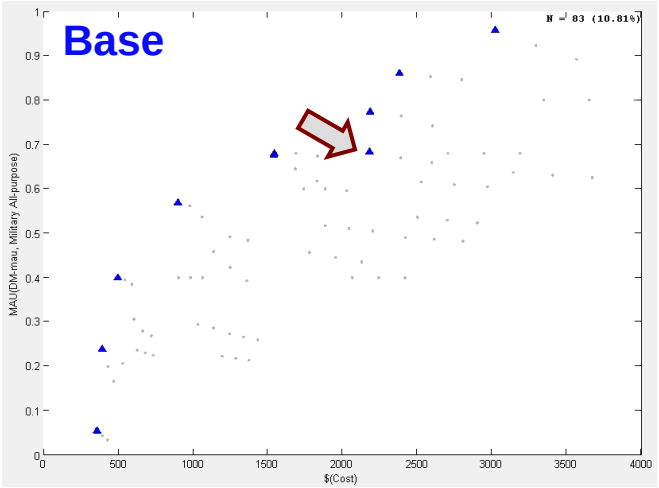
- Cat. **D** is efficient only under binary speed model
  - Nuclear, with the maximum fuel mass in the tradespace
  - Are only efficient when *not* penalized for the extra mass

Different designs may appear efficient depending on how the attributes of interest to the stakeholder are operationalized

Modeling Artifacts

- Not all conclusions from cross-model comparison have physically-intuitive explanations
- Also calls out modeling artifacts
  - "Fake" insights that might not be caught by a single implementation
- Cat. **E** is a single design that is efficient only in the base model implementation
  - Sits in a concave region of the Pareto front, nearly dominated
  - Efficiency highly sensitive to any model changes, base is a corner case

	Design ID (Aluminum)	Model 1	Model 2	Model 3	Model 4	Design ID (Carbon)	Model 3	Model 4
E	127	√				511		



This one is visually clear – but what about more subtle modeling artifacts?

Artifacts Appear via Combinations of Model Choices

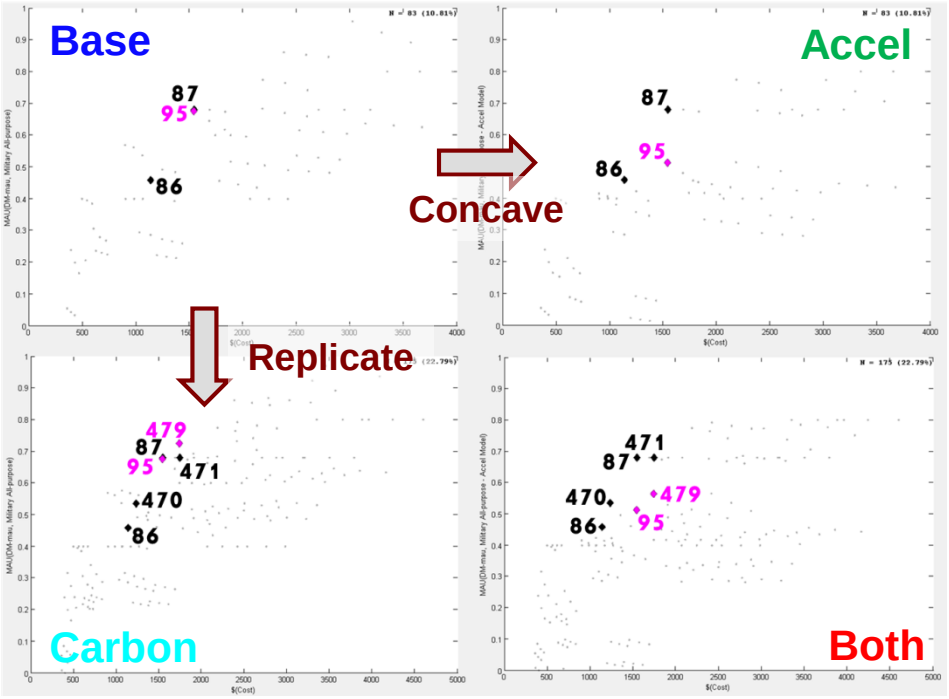
- Cat. **F** is always efficient except for implementation #4
  - What would make a design efficient across multiple speed and material models but not when they are combined?

F

Design ID (Aluminum)	Model 1	Model 2	Model 3	Model 4	Design ID (Carbon)	Model 3	Model 4
95	√	√	√		479	√	

- Adding acceleration model creates concave pattern with designs in other categories (B + C)
- Adding new material type replicates the existing pattern, slightly offset
- Combined, the concavities overlap and dominate the interior points (category F)

This choice is worse than may originally appear



“Promising” Design Sorting

- All four “promising” designs from the Value Model Trading study (valid for MAU) are sorted into different categories across the evaluative models

	1	11	63	95	127	128
$(Cost)	96.876	105.24	900	1540	2180	3020
$(Cost)	96.876	105.24	900	1540	2180	3020
$(Cost)	96.876	105.24	900	1540	2180	3020
$(Cost)	96.876	105.24	900	1540	2180	3020
$(Cost)	96.876	105.24	900	1540	2180	3020
dv(DesignforChange)	0	0	0	0	0	0
dv(DesignID)	1	11	63	95	127	128
dv(PayloadMass)	300	300	1000	3000	5000	5000
dv(PropMass)	30	300	10000	10000	10000	30000
dv(PropType)	1	2	4	4	4	4
x(Capability)	300	300	1000	3000	5000	5000
x(DeltaV)	142.608	1697.46	16149.6	10984.1	8387.01	14948.9
x(ResponseTime)	1	1	1	1	1	1
x(DeltaV)	142.608	1697.46	16149.6	10984.1	8387.01	14948.9
iv(BaseMass)	0	0	1000	1000	1000	1000
iv(DryMass)	603.6	639	5000	9000	13000	17000
iv(SpecificImpulse)	300	450	1500	1500	1500	1500
iv(MassFrac)	0.12	0.13	0.2	0.2	0.2	0.2
iv(WetMass)	633.6	939	15000	19000	23000	[illegible]
MAU(DM-mau, Military All-purpose)	NaN	NaN	0.5692	0.67607	0.68376	0.[illegible]
u(DM-mau, x(Capability))	NaN	NaN	0	0.7	1	1
u(DM-mau, x(DeltaV))	NaN	NaN	0.92299	0.49017	0.[illegible]	[illegible]
u(DM-mau, x(ResponseTime))	1	1	1	1	1	1
AHPbenefit(DM-ahp, AHP Military All-purpose)	0.20114	0.2161	0.4147	0.53522	[illegible]	[illegible]
CBAbenefit(DM-cba, CBA Military All-purpose)	4500	4702.108	7076.5553	7746.3006	78[illegible]	[illegible]

Promising Designs (3 out of 4 Pareto)	Pareto for	Invalid for
1	AHP, CBA, MOE	MAU
11	AHP, CBA, MOE	MAU
63	MAU, CBA, MOE	
95	MAU, CBA, MOE	
127	MAU, CBA, MOE	
128	MAU, CBA, MOE	

“Promising” Design ID	Category
63	A
95	F
127	E
128	D

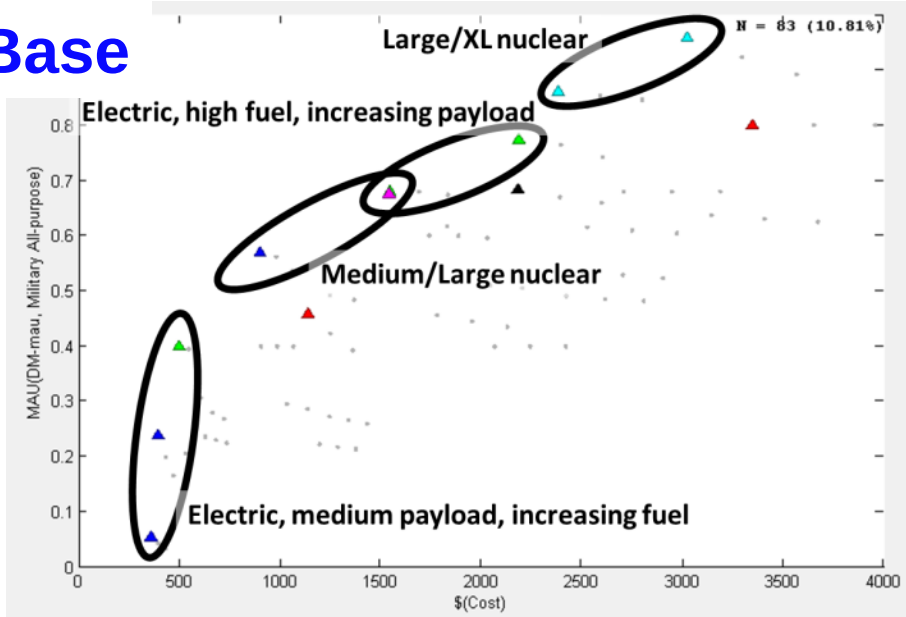
Artifact?

Consistent “goodness” across value models does not necessarily carry over across evaluative models in the same way

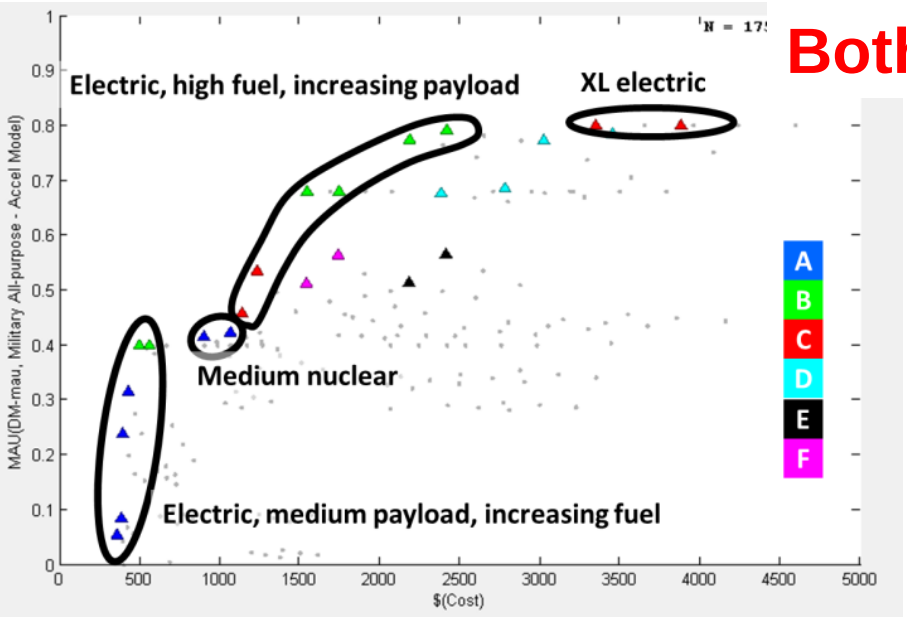
Design Variable Patterns

- Patterns of decisions in different regions of the tradespace change in different implementations
  - Low Cost: consistent
  - Mid Cost: #4 favors B and C over F with “bumped out” convex region
  - High Cost: #4 has extended range (due to carbon cost) and C dominates D

Base



Both



Electric designs (B/C) are more robust to model changes than nuclear (D/E/F)

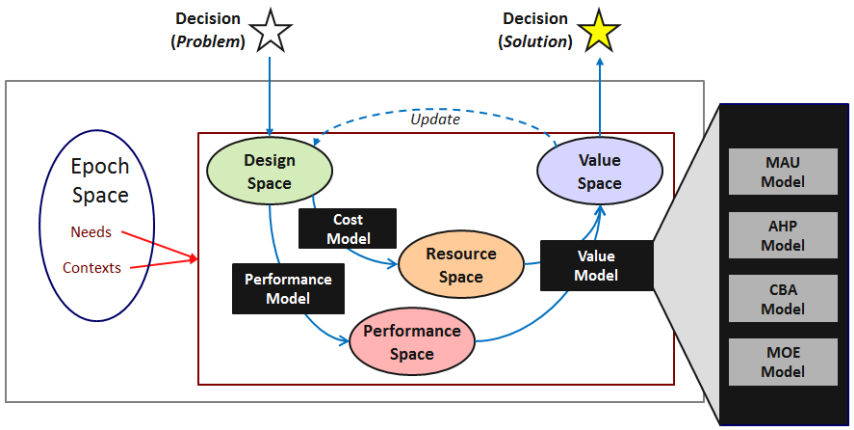
Discussion

As with value models, different "best" solutions are a natural result of evaluative model choice

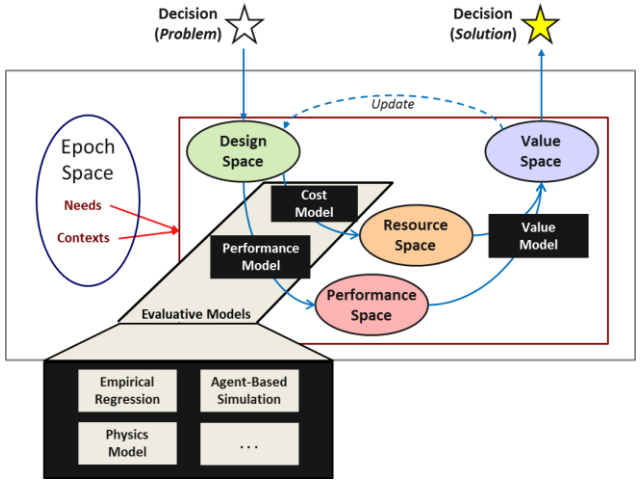
- More intuitive but no less impactful
- Can identify designs that are robust to uncertainty in model parameters or operationalization
  - Endemic to early-concept design, where "fit for purpose" not clear
  - Each implementation results in different Pareto sets with variable amounts of overlap
- Capture modeling artifacts difficult to see in a single model implementation
  - Sensitivity to change highlights unintuitive results
- Learn about how models favor some choices (unintentionally?)
  - e.g. acceleration model favors electric propulsion
  - More nuanced understanding than "optimal or not"

Conclusions

Value Model Trading



Evaluative Model Trading



- Model choice occurs throughout the design-value loop
- Model choice impacts the relative attractiveness of different choices
- Comparing models develops further insight into case
  - Use to build confidence in knowledge / trust in model
  - Use to select a single model to move forward with (e.g. optimization), or groups for ensemble analyses
  - Use to identify alternatives robust to model choice

Explicitly considering choice of models supports resilient decisions

Thank you for your attention!

Questions?

This work is supported, in whole or in part, by the U.S. Department of Defense through the Systems Engineering Research Center (SERC) under Contract HQ0034-13-D-0004. SERC is a federally funded University Affiliated Research Center managed by Stevens Institute of Technology. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the United States Department of Defense.