2016 Conference on Systems Engineering Research

Interactive evaluative model trading for resilient systems decisions

Adam M. Ross[a]*, Matthew E. Fitzgerald[a], and Donna H. Rhodes[a]

[a]*Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA 02139, USA*

Abstract

The Interactive Model-Centric Systems Engineering research effort is interested in developing knowledge necessary to leverage the increasing involvement of computational models in system design. One of the key means of leveraging a model-centric environment is the trading of models, which can reveal insights about the system that are difficult or impossible to see when considering only a single model. Prior work has demonstrated this technique on the value models used to determine the "goodness" of alternatives based on their performance and cost attributes. This paper extends the model trading paradigm to evaluative models: those that calculate the attributes themselves. The concept is demonstrated on a matching Space Tug case study with four different model implementations, each of which results in a different set of Pareto-efficient solutions. Analysis across implementations, as opposed to within a single model, reveals interesting design insights, of which some are physics-driven and others are identified as model artifacts.

Keywords: models; modeling; evaluative models; fidelity; robust solutions; tradespace exploration; resilience

1. Introduction

Models have increased in importance for engineering practitioners with the continued rise of powerful and accessible computation and, increasingly over the past few years, researchers have sought to develop standards and procedures allowing models to be leveraged more effectively[1]. Models are used to rapidly and accurately assess potential system concepts early in the design lifecycle and to explore the value tradeoffs associated with decisions designers can choose to make. Additionally, models can perform tasks that humans may struggle with, including the consideration of hypothetical scenarios that form the basis of uncertainty[2]. This paper discusses continuing research results in the area of model comparison and trading: the active comparison of different model-generated results in order to learn more about the system of interest.

* Corresponding author. Tel.: +1-617-253-7061; fax: +1-617-253-3641.
E-mail address: adamross@mit.edu

1.1. Motivation

Fig. 1 illustrates a conceptual framework for the decision process in early system design, including the relationships between the various models that are used to support the decision maker. The general flow involves the creation of a design space suited to the problem in which each design is evaluated using a set of evaluative models to determine its performance and cost (i.e., resources required) with respect to a set of given contextual factors. Those performance and resource attributes are then fed into a value model in order to assess the "goodness" of each alternative, which is the key decision-making criterion.

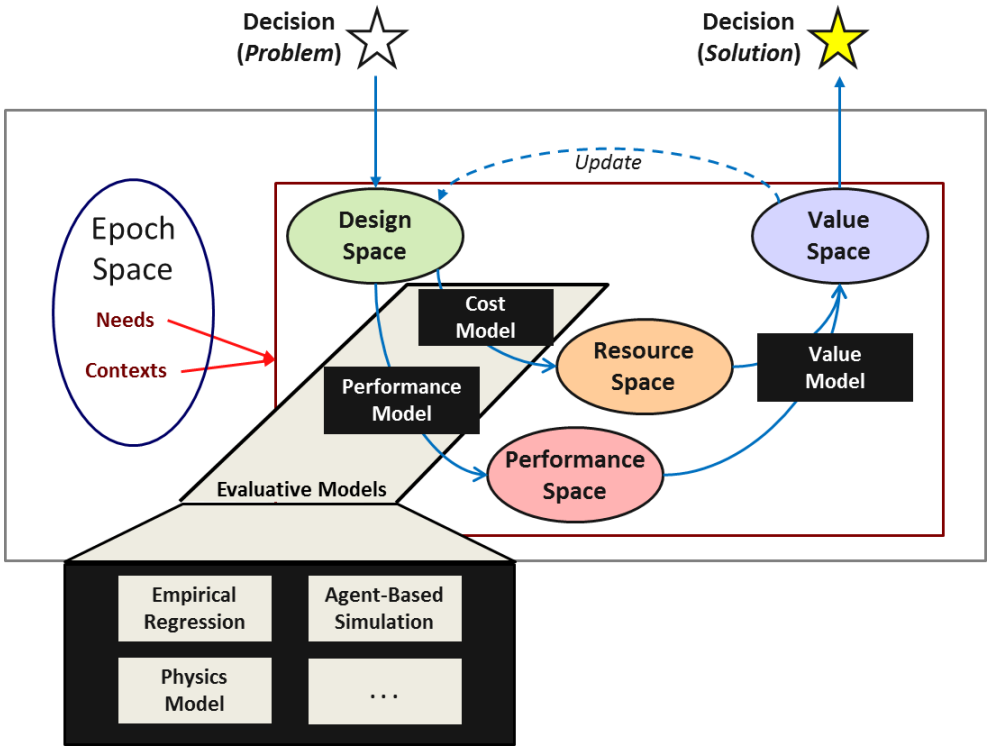

Fig. 1. Role of key models in system decision making, with example evaluative models called out.

The role and impact of models on the design process is a core interest of the Interactive Model-Centric Systems Engineering community. In our previous paper, we expanded the earlier concept of interactively refining a value model[3] into the potential use of value model trading to support the decision making process[4]. Specifically, the ability to compare the tradespace as it exists under multiple value models was shown to be a powerful means for building trust in the model results, particularly when a decision maker may be unsure how to mathematically represent their needs early in the system lifecycle. This paper addresses a similar model trading concept, this time for the evaluative models.

Evaluative models come in a plethora of different forms, too many to list exhaustively. A few common examples of evaluative models are provided in Fig. 1. Depending on the system in question, different types of evaluative models will be appropriate and/or available. Commonly discussed characteristics of evaluative models include fidelity (or the similar concepts of accuracy and precision) and computational costs (among others such as purpose and credibility[5]). When choosing models based on fidelity and computational cost, there is often a tradeoff, so the "best" model for a given task may be subjective. This choice may also require the consideration of the confidence the decision maker has in each model, which may not be solely determined by its fidelity. This best-model-calculus and associated tradeoffs are interesting topics and are likely what most readers will think of first when hearing the phrase "evaluative model trading" (e.g., determining a model's "fit for purpose" and associated verification,

validation, and accreditation (VV&A) activities[6]). However, this paper is not concerned with *selecting* the best evaluative model from a set of choices, but rather *leveraging* the ability to use multiple evaluative models in order to support the decision process. That is, instead of expending efforts to find the "right" model, what can be determined by leveraging multiple different models in order to garner potentially novel insights, especially when "fit for purpose" may be unclear early in the system lifecycle?

Why might engineers be interested in using multiple evaluative models? On a basic level, running multiple evaluative models and comparing their results can support cross-validation of each model and increase decision maker confidence in their results. Of particular interest to this research is the use of models to support early concept decision making, which may require measuring the expected performance of new or emerging technologies that have yet to be built or tested. In this case, it is possible that no evaluative models are truly validated, leading to a situation similar to that of the value model trading problem, but instead of having no "ground truth" to validate against (since value is subjective), the designers simply *don't know* what that ground truth is. As a result, searching for alternatives that are robust to the unknown accuracy or precision of the models is a powerful use of multiple models. The following case demonstrates this concept on a simple example.

2. Demonstration of evaluative model trading: Space Tug

In order to demonstrate the effects of trading evaluative models and the insights that can be gained by doing so, we will return to the Space Tug case used in the prior demonstration of value model trading[4]. The generic mission is therefore the same but the key questions have changed:

> A decision maker has a budget for an orbital transfer vehicle (a.k.a. "Space Tug") and knows what he wants (in terms of attributes of goodness of a system). However, he is aware that Space Tugs are a developing technology, and the models used to evaluate them are not 100% validated. He therefore wants to explore a variety of model implementations in order to understand the following:
>
> 1. How do changes in the model impact the apparent "best" solutions?
> 2. Are there system designs that are robust to changes in the model?
> 3. What patterns in the performance space are driven by model artifacts?

In addition to these questions, a decision maker may also be interested in exploring the tradeoffs between model fidelity, confidence in decisions, and computational effort. However in this simple example, the models have effectively zero computational cost and therefore we will focus on the implications to the system design only.

2.1. Models used in the case

The value model is held constant in this demonstration, using the multi-attribute utility model described in the previous study. Four different evaluative model implementations were tested and compared. These implementations were not created to teach specific lessons; all insights gained from their analysis were emergent.

2.1.1. Implementation #1

The Space Tug's default evaluative implementation is a combination of models[7] including a manipulator capability lookup table, a binary fast/slow speed assignment, linear models for mass and cost, and the rocket equation. There are four design variables (input elements under designer's control) leading to 384 different design alternatives, and four evaluated attributes (model outputs corresponding to the value of the system). Fig. 2 shows the default model implementation and resulting benefit-cost tradespace, with the Pareto set marked with blue triangles. Only 83 of the 384 evaluated alternatives are feasible, for a yield of approximately 21.6%. This is identical to the multi-attribute utility tradespace in the value model trading study.

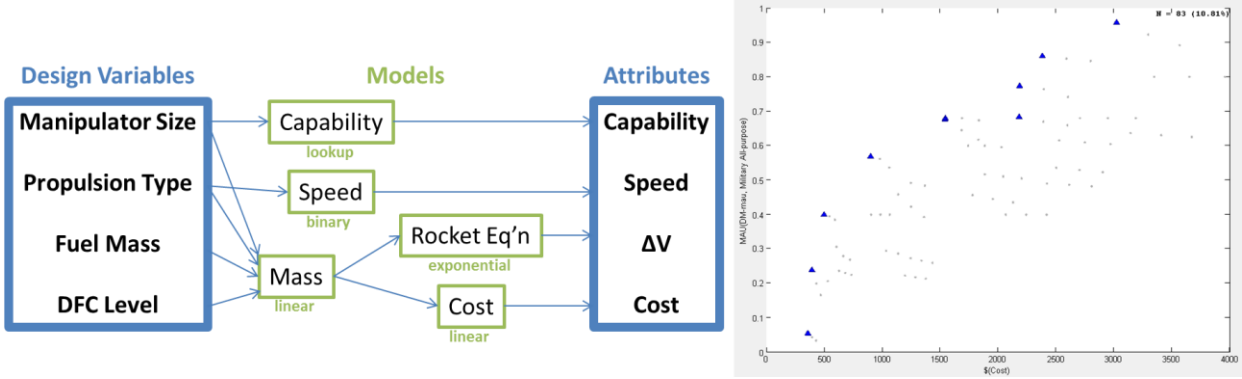

Fig. 2. Space Tug model implementation #1 and resulting tradespace.

2.1.2. Implementation #2

The second implementation is a classic fidelity upgrade. The binary speed model has been replaced by an acceleration model based on the thrust of the chosen propulsion type and the mass of the system using the classic F=m*a formulation. This model provides a more accurate estimate for the resulting speed of the system. Note that this change does not impact the yield: this is because no alternatives were designated infeasible using the original speed model but feasible using the new model (or vice-versa). Fig. 3 shows this implementation and tradespace, with the Pareto set marked in green, left-pointing triangles.

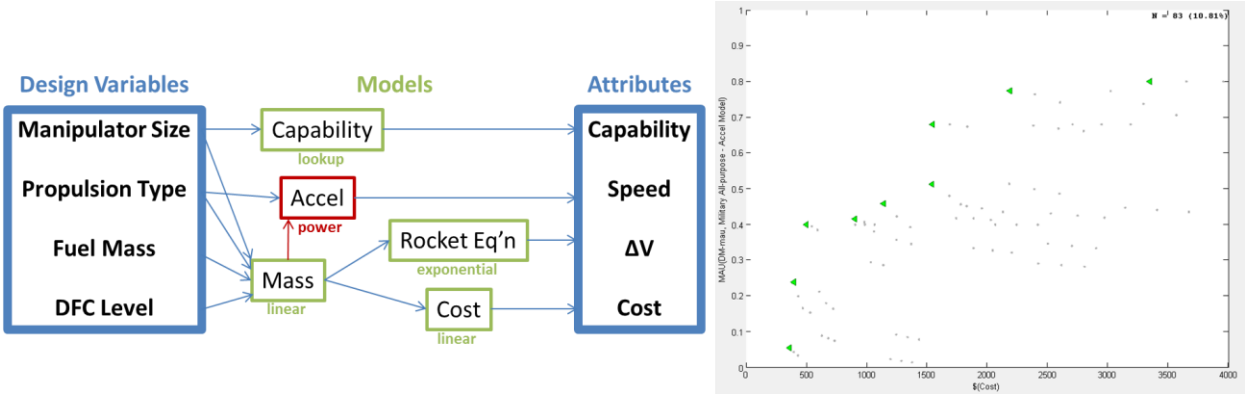

Fig. 3. Space Tug model implementation #2 and resulting tradespace.

2.1.3. Implementation #3

The third implementation is a "representational" fidelity upgrade. The mechanics of the models remain fixed but the tradespace has been expanded by taking the material comprising the support structure, previously assumed to be aluminum, and making it a design variable called *material*, accepting "aluminum" or "carbon" as choices. This allows for a more detailed representation of the system through the inclusion of other material options into the tradespace. Now the tradespace doubles in size to 768 alternatives by adding the choice of a carbon structure. The mass and cost models must accommodate this change, as carbon has a lower density but higher cost than aluminum. With this model implementation the yield *does* change, increasing to approximately 23% thanks to designs that, while infeasible with aluminum structure due to insufficient delta-V, are feasible with the lighter carbon structure. Fig. 4 shows this implementation and tradespace, with the Pareto set marked in cyan, down-pointing triangles.

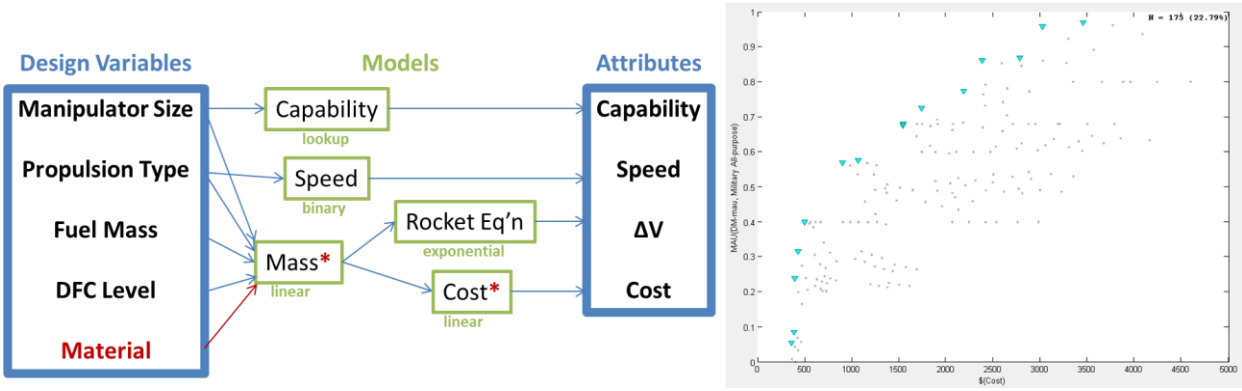

Fig. 4. Space Tug model implementation #3 and resulting tradespace.

2.1.4. Implementation #4

The final implementation combines the model changes in #2 and #3, having a more detailed acceleration model *and* the expanded tradespace. Evaluative models can be hybrid in nature, allowing for mixing and matching sub-models as desired. Fig. 5 shows this implementation and tradespace, with the Pareto set marked in red, right-pointing triangles.

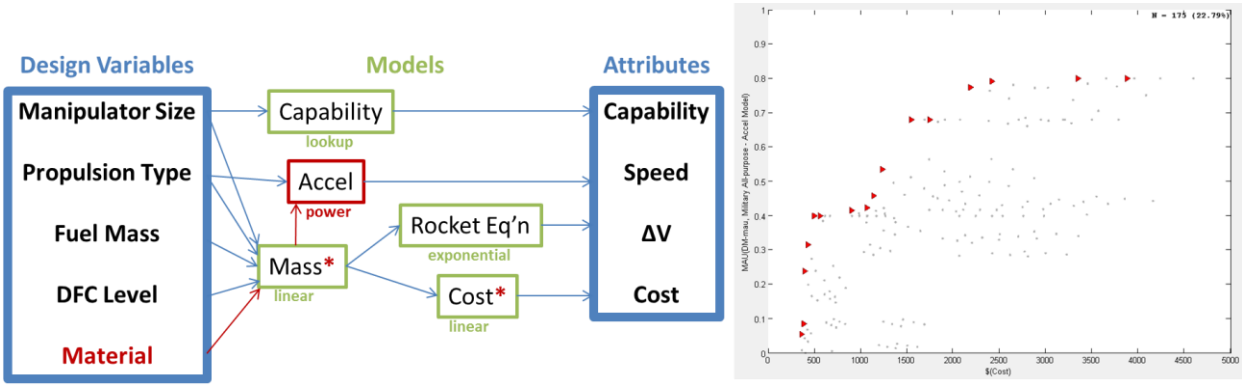

Fig. 5. Space Tug model implementation #4 and resulting tradespace.

3. Results

3.1.1. Comparisons via Pareto sets

Fig. 6 shows all four tradespaces with all four Pareto sets marked. On inspection, there appears to be more overlap between the sets for this evaluative model trading study than for the prior value model trading study. This is a positive result: while value models can drastically reorder the value of different alternatives due to differing subjective interpretations, it would be indicative of poor "fit for purpose" if the evaluative models presented dramatically different estimates of the key performance attributes. In particular, there is considerable agreement in the Pareto sets in the low-cost domain of the tradespace, suggesting that the differences between these implementations are experienced mostly in larger, more expensive systems. Fuzzy Pareto set analysis was conducted as it was in the prior paper, but is omitted here for two reasons: (1) designs that are jointly efficient between implementations already exist at zero fuzziness and (2) the resulting patterns between implementations are extremely similar to those identifiable in the basic Pareto sets, but more difficult to "see" when plotted on the

tradespace. For the record, considering a 1% fuzzy Pareto set approximately doubles the number of designs under consideration in each model implementation, while 5% approximately triples it.

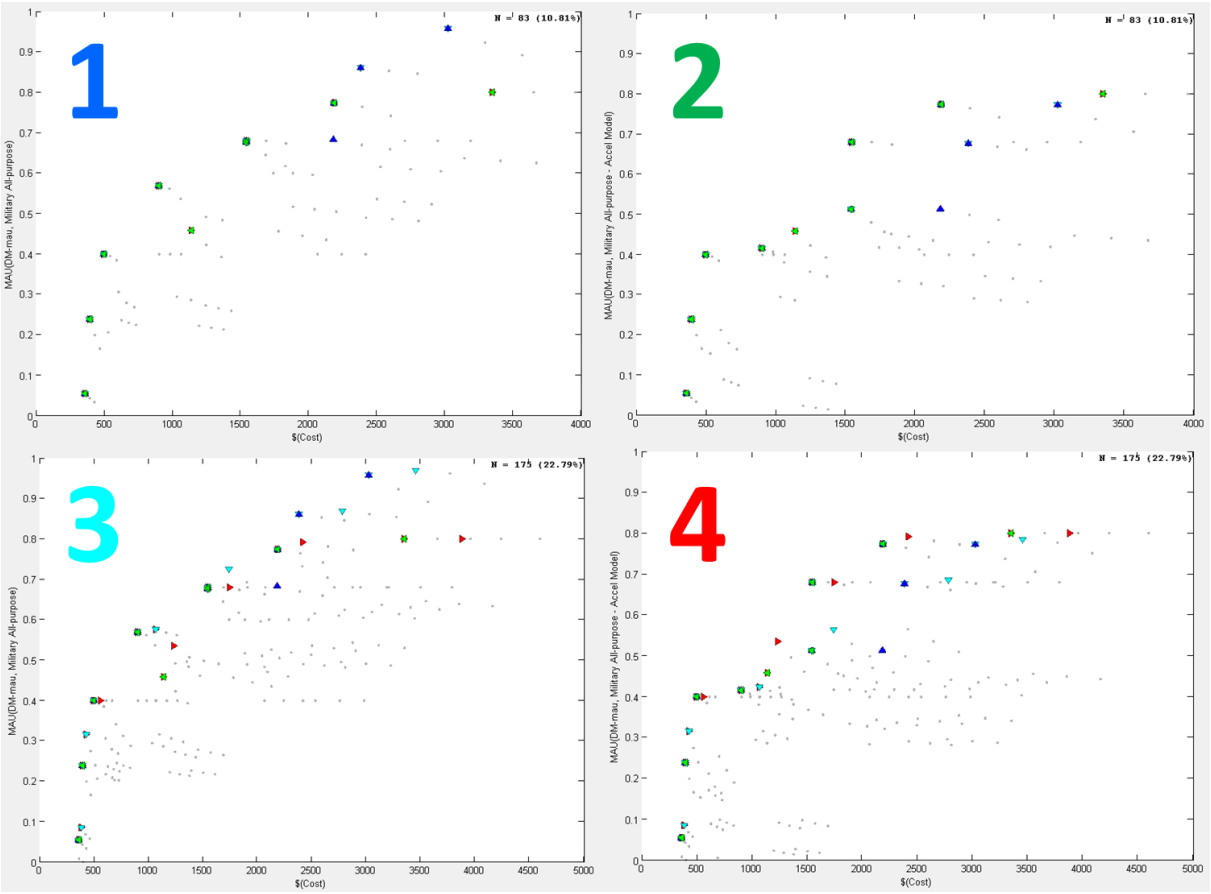

Fig. 6. The tradespaces for all four implementations, with all four Pareto sets marked in each

3.1.2. Joint Pareto analysis

To explore these designs in more detail, Table 1 includes all of the Pareto-efficient marked designs in the above figures and shows in which implementations they are efficient. Many designs that are efficient with aluminum material are also efficient when changing to carbon in model implementations 3 and 4, therefore the table is set up such that designs that are identical except for material are placed in the same row. From this list there are six emergent categories of efficient designs, which we will identify as A through F:

- **A.** Designs 52, 53, and 63 (and their carbon counterparts) are always efficient, in every implementation. These alternatives are robust to evaluative model trading, and are all among the low-cost solutions previously identified by inspection of the scatterplots.

- **B.** Designs 54, 87, and 119 are always efficient except for their carbon variants under the binary speed model in implementation 3. All of these alternatives have electric propulsion (categorized as "slow" by the binary model) and delta-V above or very near the maximum-utility point in the utility function (20,000 m/s). This results in no additional benefit when switching to carbon, but the additional costs are still experienced. In implementation 4

however, the reduced weight from the carbon structure leads to improved acceleration, making these designs efficient again.

Table 1. Designs, marked in gray and with a check for model implementations in which they are efficient

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

- **C.** Designs 86 and 120 are efficient in both materials but *only* under the improved speed model of implementations 2 and 4. These designs are the same as 87 and 119 except for slightly more and slightly less fuel, respectively. The higher fidelity acceleration calculation results in slightly different "sweetspots" in the tradeoff between speed and delta-V driven by fuel mass.

- **D.** On the other hand, designs 96 and 128 are efficient in both materials but only under the *original* speed model in implementations 1 and 3. These designs both have the maximum fuel mass enumerated in the tradespace. This extra mass does not penalize speed under the original model, allowing these designs to be efficient; however, the switch to the acceleration model lowers their value and removes them from the Pareto front.

- **E.** The above categories have physically-intuitive explanations for their behavior, but some insights from model trading do not. Trading models can also show that some conclusions that could be drawn from a single tradespace are nearly entirely artifacts of the model implementation. For example, design 127 is efficient *only* in the original implementation. This design is the same as 128 but with less fuel, and sits in a noticeably concave region of the Pareto front, which is surpassed by many designs when considering higher fidelity and/or alternate material models. Model trading reveals this design to be much less attractive than it originally appears.

- **F.** Perhaps the most unintuitive insight of all is that design 95 and its carbon counterpart 479 are efficient in implementations 1, 2, and 3 but *not* 4. It is not immediately apparent why a given system design would be efficient using the original models and higher fidelity models, but not when those higher fidelity models are combined, especially for a simple case such as Space Tug where most of the equations in play are linear. Fig. 7 shows the four tradespaces with 95/479 highlighted in magenta and 86/87/470/471 highlighted in black. In implementation 1 (top-left), 95 is closely co-located with 87. The shift to implementation 2 and the acceleration model (top-right) causes 95 to decrease in benefit and occupy a concave region of the Pareto front between 86 and 87. On the other hand, shifting to implementation 3 (bottom-left) nearly replicates the implementation 1 pattern with the three carbon designs, shifted to slightly more cost and benefit but such that they still overlap the original three designs. However, combining these two effects in implementation 4 (bottom-right) results in

overlapping concave patterns, with the "elbow" points in each triplet, designs 95/479, now dominated by an end point of the other triplet: 95 by 470 and 479 by 87. This illustrates the benefit of model trading for capturing deep insight into the tradespace, as this interaction indicates that, despite the inherent simplicity of this example, there is some unpredictable interplay between the speed and mass submodels that may merit further detailed analysis of their component functions.

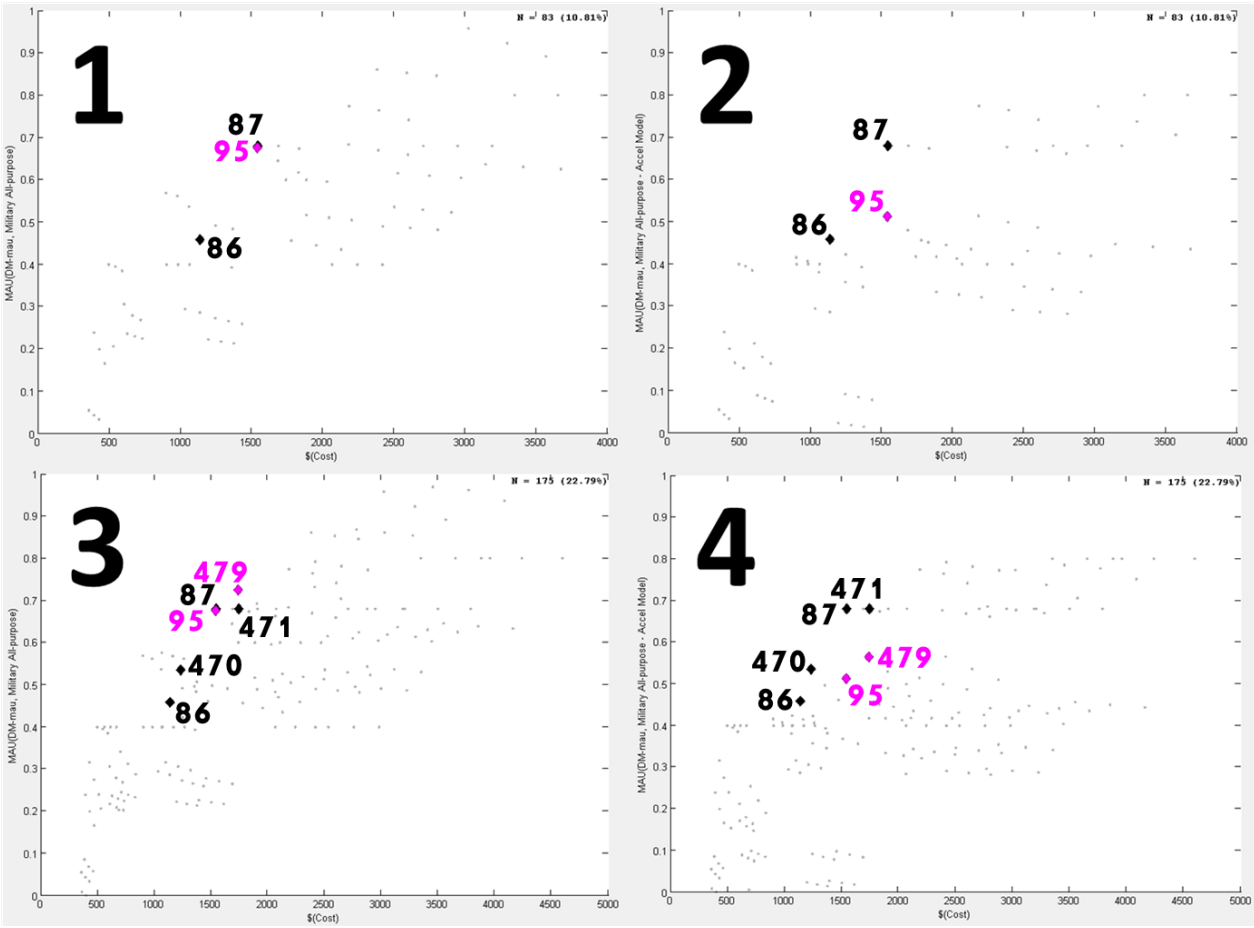

Fig. 7. The tradespaces for all four implementations, highlighting the model artifacts impacting designs 95/479 (magenta), combining to make them inefficient in implementation 4 only. Design numbers left/above of the points are on the Pareto front, right/below are inefficient.

The four "promising" designs identified by the value-model trading study (63, 95, 127, and 128) are all present in the above list, a fact that was guaranteed since the prior analysis confirmed their efficiency in the equivalent of implementation 1. However, it is interesting to note that each of these designs falls into a different category of interest across the different evaluative implementations. Combining the insights of the two studies suggests that designs 127 and 128 are not as good as originally believed, but 63 and 95 (or their carbon variants, which were not in the value-trading study) are still potentially interesting selections.

Finally, we can attempt to build some intuition for the impact of model implementation in terms of the design variables of the efficient alternatives. Fig. 8 shows implementations 1 and 4 with the different Pareto efficient designs colored by which category they belong to in Table 1. In each tradespace, regions of the Pareto front with consistent patterns are highlighted. First, note that no bipropellant or cryogenically propulsed Space Tugs are efficient: this is the result of the multi-attribute utility function, which was noted in the prior study to eliminate most

of those alternatives due to high requirements on delta-V. However, we can see here that the relative value of the remaining Pareto efficient electric and nuclear designs is sensitive to model implementation. While both tradespaces have medium payload, electric and nuclear Space Tugs in the low cost region (members of category A and the low end of B), there are two main impacts of the model changes. First, in the medium cost region, implementation 4 has a convex region that strongly favors the electric designs of categories B and C, whereas implementation 1 is nearly linear and features crossover between the electric designs and large nuclear designs of F. Second, the extra-large nuclear designs of category D dominate the high cost region in implementation 1 but switch to high-end electric designs of C in implementation 4. Overall then, it seems that the new acceleration model favors electric designs more than the original model by closing the gap on both ends – giving electric designs more credit for acceleration than a binary '0' and penalizing high-mass nuclear designs. This means that electric designs are more robust to these model trades, while the regions of the tradespace for which nuclear designs are optimal will change more dramatically depending on implementation. However, it is worth pointing out again that the reordering of these designs is less dramatic than it was for value model trading, as this type of analysis captures slight differences in efficient and nearly-efficient designs, rather than a complete redefinition of value.

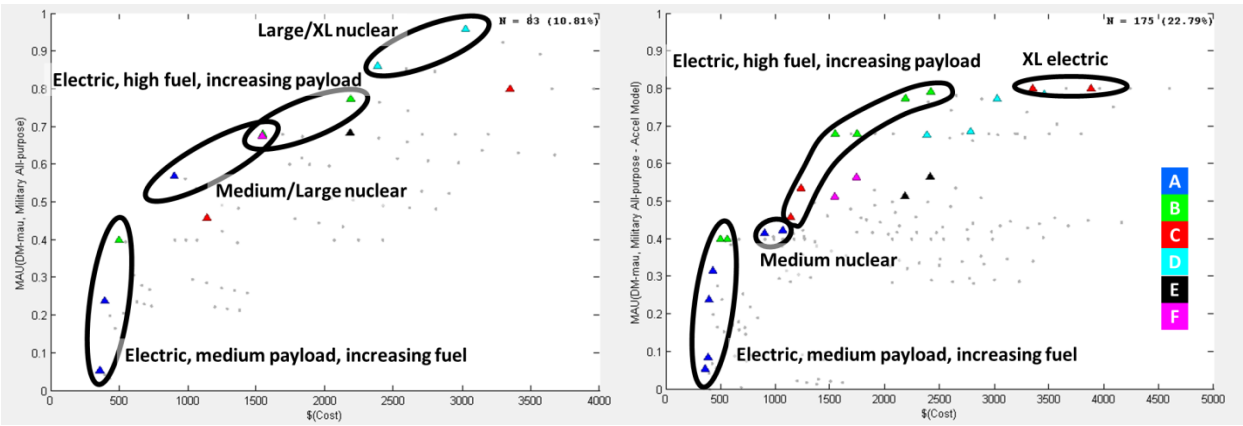

Fig. 8. The tradespaces for implementations 1 (left) and 4 (right), highlighting design variable patterns on the Pareto front. The six categories of designs in Table 1 are used as the color scheme.

4. Discussion

Designers can gather insight about their preferences from the exploration of value model trades, which can be a useful exercise given the lack of ground-truth data to support their validation. Evaluative performance and cost models can similarly be traded and explored, which may become particularly worthwhile for early-concept designs involving emerging technology too new to have established and validated supporting models. These evaluative model trades often take the form of fidelity differences in either representation (of the design) or evaluation (of the system attributes), and typically come with an associated difference in computational cost. In some cases, model trades may encompass completely different physical phenomenologies, such as when engineers must select a turbulence model within a larger fluid dynamics model – a decision that is often made with the support of experimental prototyping data to identify the model that is most accurately predicting the resulting flow. In this way, evaluative model trading can support both the identification of approximate bounds or errors on the results of model-centric engineering efforts and build designer trust and confidence in data by revealing the impacts of switching model components and the idiosyncrasies that can arise at the intersection of multiple models. The demonstration case provided here was intended to focus on the latter of these points, as building of model trust is considered a strength of the tradespace exploration paradigm. Future studies can also explore the tradeoff between model fidelity, designer confidence, and computational effort, which will require a case with a more complex model with a higher available top-end fidelity and longer resulting computation times. Additionally, the combination of both value model trading and evaluative model trading is anticipated to provide further insights into the impact of

model choice, and emergence within such models, on potentially attractive design solutions. The next phase of research will seek to combine the two into a coherent model trading framework that allows for engineers to explore the impacts of uncertainty in both domains at once. This research will also be integrated with other efforts of the developing Interactive Model-Centric Systems Engineering research effort in support of its larger goals for effective, integrated modeling and exploration environments and practices[8,9].

Acknowledgements

This material is based upon work 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.

References

1. Rhodes DH, Ross AM. Interactive Model-Centric Systems Engineering (IMCSE) Phase 1 Technical Report. SERC-2014-TR-048-1; September 2014.
2. Spero E, Avera MP, Valdez PE, Goerger SR. Tradespace exploration for the engineering of resilient systems. 12th Conf on Sys Eng Research. (CSER14). Redondo Beach, CA, March 2014.
3. Ricci N, Schaffner MA, Ross AM, Rhodes DH, Fitzgerald ME. Exploring stakeholder value models via interactive visualization. 12th Conf on Sys Eng Research. (CSER14). Redondo Beach, CA, March 2014.
4. Ross AM, Rhodes DH, and Fitzgerald ME, "Interactive Value Model Trading for Resilient Systems Decisions," 13th Conf on Sys Eng Research. (CSER15). Hoboken, NJ, March 2015.
5. NASA. Standard for Models and Simulations. NASA-STD-7009; July 2008.
6. INCOSE. Systems Engineering Handbook: A Guide for system life cycle processes and activities. 4th ed. Chapter 9.1 Modeling and Simulation. INCOSE-TP-2003-002-04. 2015.
7. McManus H. Schuman T. Understanding the orbital transfer vehicle trade space. AIAA Space 2003. Long Beach, CA, September 2003.
8. Reymondet L, Rhodes DH, and Ross AM, "Considerations for Model Curation in Model-Centric Systems Engineering," IEEE Syscon 2016, Orlando, FL, April 2016.
9. Rhodes DH, and Ross AM, "A Vision for Human-Model Interaction in Model-Centric Systems Engineering," INCOSE Int'l Symp 2016, Edinburgh, UK, July 2016.