Blind Validation Study of Parametric Cost Estimation Tool SEER-H for NASA Space Missions

Paul D. Friz*[1], Serhat Hosder[2], Benjamin B. Leser[3], Benjamin C. Towle[4]

Abstract

One of the primary parametric cost modeling tools used by NASA and the aerospace industry to estimate the development and production cost of future spacecraft hardware is SEER-H by Galorath. To date, no independent validation of this tool for cost estimation of space missions has been reported in the literature. In the present validation study, cost estimators used SEER-H to estimate the cost of twelve different past NASA science missions. The estimators were prevented from knowing the actual cost of the missions in an effort to minimize cognitive biases. The point estimates of SEER-H had an average error of 23%, median error of -0.3%, and a standard deviation of 43%. Nine of the twelve mission's actual costs fell within the 80% confidence interval of SEER's probabilistic estimates. Several factors independent of SEER that may have affected the accuracy of the results have been identified and are discussed; these include: uncertainty in the technical data used for the estimates, the methods used to estimate uncertainty in spacecraft component mass and numbers of prototypes, and the experience of the estimators.

Keywords: SEER-H, parametric costing tools, validation, cost estimation

---

*Corresponding author
Email Address: Paul.D.Friz@nasa.gov
[1] NASA Langley Research Center
[2] Missouri University of Science and Technology
[3] University of Michigan-Ann Arbor
[4] University of Michigan-Dearborn

Preprint submitted to Acta Astronautica *May 8, 2019*

1. Introduction

Parametric cost estimation tools are commonly used both at NASA and in industry to predict life cycle costs of space systems[1, 2, 3]. Parametric costing tools are software packages that use regression analysis to relate quantitative technical parameters and qualitative nontechnical parameters to cost[4]. Examples of quantitative technical parameters include mass, volume, and part count. Nontechnical parameters are equally important but are difficult to measure and or subjective items such as complexity, level of new design, and team experience. Parametric costing tools allow users to quickly estimate the costs of a mission before detailed designs have been completed. Typically, they are used to estimate the development and production costs of hardware early in the project life-cycle, including mission conceptstudies (Pre-Phase A), concept development (Phase A), and preliminary design (Phase B). Parametric costing tools are also used to perform trade studies and predict costs of design variations and are frequently used to generate independent cost estimates to evaluate mission proposals.

One of the primary costing tools used by NASA to evaluate mission, spacecraft, and instrument proposals is SEER Hardware Estimation and Life Cycle Cost Analysis (SEER-H) by Galorath Inc., henceforth referred to as simply SEER[5]. To date, no independent validation of SEER for estimating the cost of space missions have been reported in the literature. Galorath performed their own internal validation studies of SEER, and reported the results of these studies without details on the methods used to perform the validations[6]. The primary goal of the study performed by Galorath appears to have been to develop standard inputs and assumptions to model various space hardware rather than to validate the model against actual costs. These input assumptions are documented in the SEER-H Space Guidance document, an extremely valuable resource for anyone using SEER to estimate the cost of a space system[7].

Galorath's validation study of SEER consisted of fifteen case studies of various robotic NASA spacecraft, and included modeling 46 instruments[6]. Both

Galorath's study and the present study make use of the Electro-Optical System (EOS) and Integrated Circuit (IC) plug-ins for SEER. Galorath provides little information on how the study was conducted other than the names of the missions used and that "A standardized modeling approach was utilized, which formed the basis of a Space Guidance document to be released in the future." The SEER Space Guidance document has since been released and was used to form the SEER estimates in the present study[7]. Of the fifteen missions in the SEER study, seven were Discovery, six were Explorer, and two were New Frontiers. Discovery class missions are low-cost solar system exploration missions typically cost capped at around $300 million. Explorer class missions are even lower cost and are typically capped at $200 million. New Frontiers missions are larger than Discovery and Explorer missions, but are not as expensive as flagship missions. New Frontiers missions typically are cost capped under $1 billion. Galorath's study found that over the fifteen missions SEER's average error in predicting cost was -1% with a standard deviation of 19%. Therefore, roughly 68% of all SEER estimates will be from -20% to +18% of the actual costs and 95% of estimates will be from -39% to +37%, assuming the mean values of SEER estimates follow a Gaussian distribution. It is not know whether or not the estimators in Galorath's study knew the mission costs before performing their estimates so that they could provide guidance to customers on how to adjust inputs, or if they conducted the study in a blind manner to validate their models.

There are several important differences between the present study and the Galorath study. It appears the primary goal of the Galorath study was to determine a standard set of inputs and settings which would produce the most accurate results when modeling the cost of a space mission. The goal of the present study was to independently assess the accuracy of SEER in an environment that matches that of an independent cost estimate as closely as possible. The Galorath study used technical data about the spacecraft and payload designs from the end of the mission; whereas the present study used technical data from each mission's Critical Design Review (CDR). In other words, the

Galorath study modeled the spacecraft as they were built, while the present study modeled how the spacecraft were expected to be built at the time of CDR. As a result, errors in the cost estimates of the Galorath study are primarily due to the uncertainty introduced SEER and how the hardware is modeled in SEER, whereas the present study also includes uncertainty in the design of the spacecraft at CDR. Thus, the Galorath study shows how accurate SEER can be, whereas the present study endeavors to show how accurate SEER is likely to be when used in its typical use case. This required that the estimators have no prior information of the mission cost, so that cognitive biases could be minimized.

The methodology of the present study is further described in Sec. 2. Section 3 presents the results of the present study, Sec. 4 discusses factors which may have affected the results of the present study that are independent of SEER, and Sec. 5 provides the conclusions of the present study.

2. Methodology

For the present study, twelve missions were selected from the One NASA Cost Engineering (ONCE) database. ONCE is a database maintained by NASA, which stores technical and managerial documents presented at major design reviews for NASA missions. These documents include the Master Equipment List (MEL) for each spacecraft, technical descriptions of the spacecraft components, and records of how much money was spent on the mission. This information is captured in documents know as Cost Analysis Data Requirements (CADRes). CADRes are a set of three documents recording:

1. An overview of the mission.
2. A mass and power breakdown.
3. A cost breakdown by subsystem and year.

CADRes are typically generated at each major review for a mission; such as, the Mission Concept Review (MCR), Critical Design Review (CDR), Launch Readiness Review (LRR), End of Mission (EOM), and potentially others. The present study used the technical data, MELs, hardware descriptions etc., from the CDR to generate cost estimates and compared them with their actual costs at either LRR or EOM, depending on which data set was available. This differs from Galorath's study, where LRR/EOM technical data was used to generate estimates that were compared to the same LRR/EOM costs. The present study used the data from CDR in part because the documents produced at CDR typically go into significantly more technical detail than the documents produced at other phases of the mission. Using the information available at CDR as opposed to LRR/EOM is also a more realistic test of the capabilities of SEER since no estimator will know the final design when producing an independent cost estimate. Using CDR data was also done to review the standard mass margin estimation assumptions discussed in Sec. 4.1.

The present study was conducted in the following manner. Cost estimator A selected the missions and removed all references to cost from the CADREs and supporting technical documents. These cleansed documents were then given to estimators B and C who used them to produced SEER cost estimates. Estimators B and C were blind to (i.e., unaware of) the actual mission costs in order to prevent cognitive biases, such as anchoring, from influencing their cost estimates. Anchoring is a cognitive bias that can cause an individual to rely too heavily on an initial piece of information when making decisions. Estimator A was an advanced SEER user and experienced cost estimator. Estimators B and C were novice SEER users, but were trained for several weeks before the start of the study.

The cost of every mission in this study was broken down into the standard NASA space mission WBS shown in Fig. 1. SEER is not capable of estimating the cost of every WBS element. SEER can only estimate the costs associated with Project Management (PM), Systems Engineering (SE), Safety and Mission Assurance (S&MA), Payload, Spacecraft Bus, and Systems Integration and Test (IAT), which correspond to the standard NASA WBS items 1-3, 5, 6, and 10, respectively. Estimates of WBS 4, 7-9, and 11: Science/Technology, Mission Operations, Launch Vehicle, Ground Systems, and Education and Public Out-

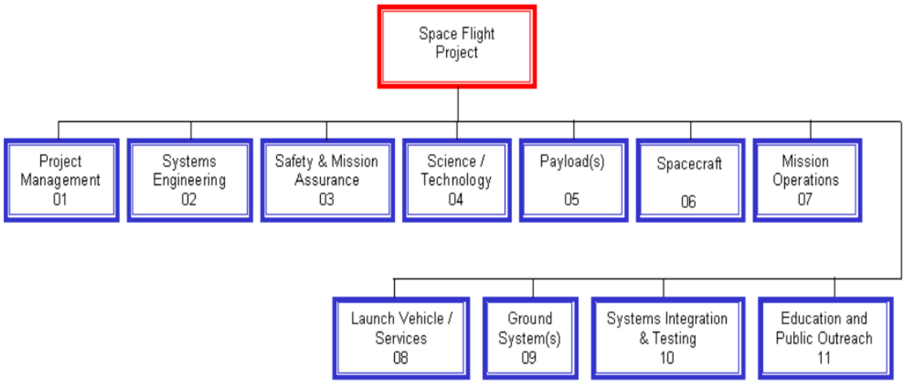


Figure 1: Standard NASA space mission WBS

reach respectively, were not included in this study. Additionally, no software costs were included in this study. The costs of the Spacecraft Bus were broken down into the following subsystems: structures, thermal control, propulsion, Guidance, Navigation, and Control (GN&C), communications, electrical power, harness, and Command, and Data Handling (C&DH). The majority, but not all, of the missions have their costs broken down in this manner, with the exceptions being attributed to data limitations. The costs included in this study are only those in mission phases B (Preliminary Design & Technology Completion), C (Final Design & Fabrication), and D (System Assembly, Integration & Test).

To ensure that proper comparisons were being made with the actual mission costs all cost estimates were converted to real-year dollars using the 2017 NASA New Start Inflation Index. As the costs of various payloads/subsystems are sometimes sensitive/proprietary, all costs presented in the present paper have been normalized.

There are three main factors that determine the accuracy of a spacecraft hardware cost estimate using parametric cost estimating tools:

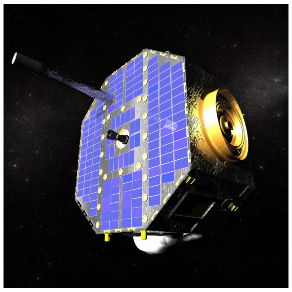
1. The accuracy and precision of the tool used to model the cost,
2. The quality and quantity of the technical data describing the hardware,
3. The knowledge, experience, and skill of the cost estimator using the cost modeling tool and evaluating the data.

The goal of the present study is to test the first factor, while minimizing the contributions of the second and third factors. Both the present study, and the validation study performed by Galorath Inc. use CADRes and supporting data from ONCE. However, for many missions there are a large number of technical documents with varying detail so there may be some slight inconsistencies between the various documents used to build the estimates. It is worth noting that although the CADRes contain significant quantities of data they do not contain every single piece required to complete a parametric cost estimate. Inevitably, the estimators were required to interpret limited data or make assumptions where data was missing.

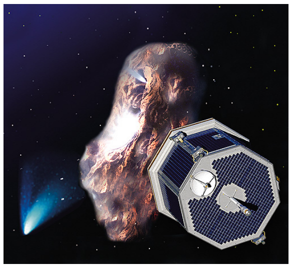
2.1. Missions Selected for the Present Study

The missions for the present study were selected from recent robotic NASA science missions which have complete CADRes on the NASA ONCE Database and sufficient supporting technical documentation to build a credible estimate. Figure 2 shows the missions selected for the present study, which include: IBEX, CONTOUR, WISE, New Horizons, MESSENGER, GRAIL, Deep Impact, MAVEN, Dawn, Kepler, SMAP, and Juno. Table 1 summarizes the class and destination of each of the missions included in the present study.

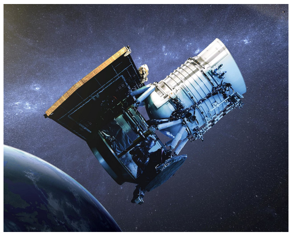
This selection represents a wide variety of robotic spacecraft types including earth orbiting, planetary, and space telescopes. The Interstellar Boundary EXplorer (IBEX) was a small Explorer class mission in earth orbit designed to study the interaction of solar wind and interstellar medium[8]. COmet Nucleus TOUR (CONTOUR) was a Discovery class mission managed by the Johns Hopkins Applied Physics Lab (APL) which was to flyby three comets[9]. Unfortunately, shortly after initiating the burn towards its first comet, the heat generated by the solid rocket weakened the aluminum support structures holding the motor in place causing them to fail. As a result the spacecraft was lost[10]. Wide-field Infrared Survey Explorer(WISE) was a small (400 cm) space telescope in low earth orbit which imaged the entire sky in multiple infrared bands. The WISE mission was an Explorer class mission managed by the NASA Jet Propulsion



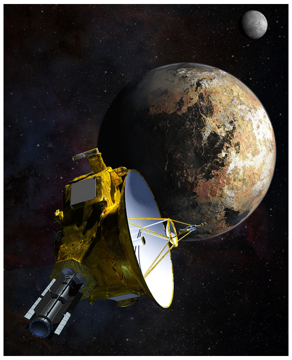
(a) IBEX



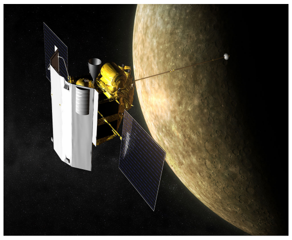
(b) CONTOUR



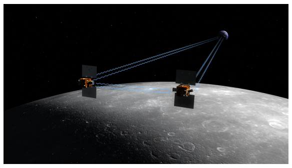
(c) WISE



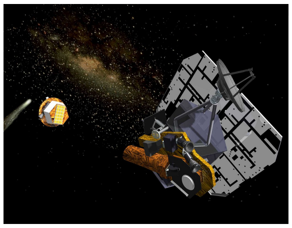
(d) New Horizons



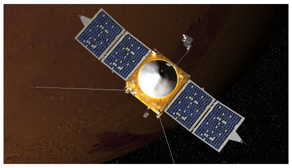
(e) MESSENGER



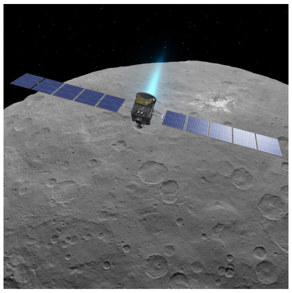
(f) GRAIL



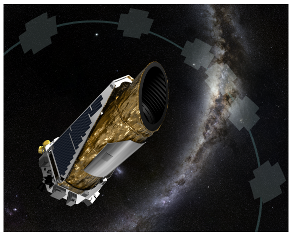
(g) Deep Impact



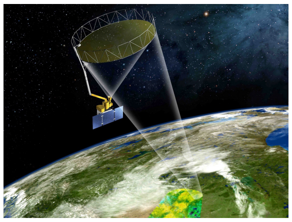
(h) MAVEN



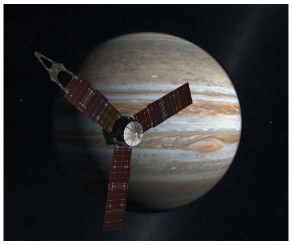
(i) Dawn



(j) Kepler



(k) SMAP



(l) Juno

Figure 2: Spacecraft included in this study.

Mission	**Class**	**Destination**
IBEX	Explorer	Earth Orbit
CONTOUR	Discovery	Three Comets
WISE	Explorer	Earth Orbit
New Horizons	New Frontiers	Pluto
MESSENGER	Discovery	Mercury
GRAIL	Discovery	The Moon
Deep Impact	Discovery	Comet 9P Tempel
MAVEN	Discovery	Mars
Dawn	Discovery	Vesta and Ceres
Kepler	Discovery	Earth Orbit
SMAP	Earth Science	Earth Orbit
Juno	New Frontiers	Jupiter

Table 1: Summary of missions included in the present study.

Laboratory (JPL)[11]. New Horizons was a New Frontiers mission managed by APL, the spacecraft was the first to flyby the planet Pluto. MErcury Surface, Space ENvironment, GEochemistry, and Ranging (MESSENGER) was a Discovery class mission and was the first spacecraft to orbit the planet Mercury[12]. Gravity Recovery and Interior Laboratory (GRAIL) was a Discovery class mission managed by JPL made up of two small spacecraft orbiting and mapping the gravitational field of the moon to determine its interior structure[13]. Deep Impact was a PI lead Discovery class mission managed by JPL. The mission was made up of two spacecraft, one which impacted comet Tempel 1 and a second which observed the impact[14]. Mars Atmosphere and Volatile EvolutioN (MAVEN) was a PI lead mission from the Mars Scout Program (The Mars Scout program was later rolled into the Discovery program). The PI was at University of Colorado Boulders's Laboratory for Atmospheric and Space Physics (LASP) while the project was managed by Goddard Spaceflight Center (GSFC). MAVEN's primary objectives were to explore the interaction of the Sun and the solar wind with Mars's magnetosphere and atmosphere to learn more about how Mars lost much of its atmosphere and surface water[15]. Dawn was a PI lead Discovery class mission managed by JPL with significant contributions from the space agencies of Italy, Germany, France and the Netherlands. The spacecraft orbited the planet Ceres, and the asteroid Vesta, and was the first to orbit two extraterrestrial bodies, and was the first NASA exploration mission to utilize ion propulsion[16]. Kepler was a PI lead Discovery class mission, the PI was at NASA Ames Research Center (ARC) and the mission was managed by JPL. The spacecraft was a wide field space telescope capable of continuously monitoring 100,000 stars at once to detect transiting exoplanets[17]. Soil Moisture Active Passive (SMAP) was an earth science mission managed by JPL. The spacecraft supports a rotating 6-m diameter lightweight deployable mesh reflector which is used by the spacecrafts active radar and passive radiometer to measure global soil moisture[18]. Juno was a PI lead New Frontiers class mission to the planet Jupiter. Juno's PI was at Southwest Research Institute(SwRI) while the mission was managed at JPL. Juno's mission was to learn more about the origin, interior, atmosphere and Magnetosphere of Jupiter[19].

Not all CADRes included sufficient technical details about mission payloads, usually this occurred in cases where payload instruments were contributed, or paid for by other space agencies or universities. All the spacecraft included in this study carried multiple instruments but Dawn and New Horizons did not have enough supporting documentation to build credible estimates of their payload costs. Thus, Dawn and New Horizons payload costs have been omitted form this study.

3. Results of Blind Study

This section presents the results of the blind portion of the study where the estimators had no knowledge of the actual costs of the missions. The next section (Sec. 4) discusses these blind results but also presents some non-blind results generated after the preliminary analysis of the blind portion of this study was complete.

While the total cost of a mission is public information, often the cost of specific pieces of hardware is proprietary or otherwise sensitive information. In order to protect against any potential breach of sensitive information all results of this study are expressed as a percent error. The percent error of each estimate is calculated using Eq. 1.

$$\varepsilon = \frac{(C_E - C_A)}{C_A} \tag{1}$$

Where ε is the percent error, C_E is the estimated cost, and C_A is the actual cost from the CADRe data.

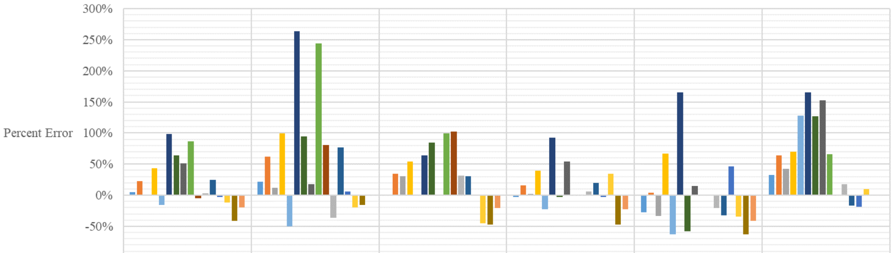
SEER outputs the cost of each spacecraft/payload component individually and outputs system level costs for groups of components know in SEER as roll-ups. System level costs are broken into project management (PM), systems engineering (SE), safety and mission assurance (S&MA), and integration and test (IAT). While SEER outputs the cost of each component individually the CADRes recording the money spent on each mission do not necessarily go into as much detail, and sometimes differ in the level of detail from mission to mission. CADRes typically record the systems level costs of NASA, the system level costs of the spacecraft bus, the cost of each spacecraft subsystem, and the total cost of each payload instrument.

In addition to presenting the error in the estimate of the total cost of each mission, the errors were broken down into the standard NASA WBS items, and the spacecraft bus estimate was further broken down by subsystem. Per the SEER space guidance, the NASA level PM, SE, and S&MA (WBS 1, 2, & 3) estimates were rolled up together while the payload total (WBS 5), spacecraft bus total (WBS 6), and IAT (WBS 10) were presented separately. In addition, the spacecraft bus total is broken down into the PM, SE, and IAT costs of the contractor building the spacecraft bus as well as the spacecraft bus subsystems: structures, thermal, propulsion Guidance, Navigation, and Control (GN&C), communications, electrical power, harness (sometimes rolled up with electrical power), and Command and Data Handling (C&DH). The mission total was defined as all of the costs that were estimated in the present study for a particular mission, namely WBS 1, 2, & 3, payload total, spacecraft bus total, and WBS 10. The mission total cost did not include Science/Technology, Mission Operations, Launch Vehicle, Ground Systems, Education and Public Outreach, software development, or any other aspects of the mission that were not included in the estimate. There were several instruments without enough supporting technical documentation to estimate their cost. Such instruments were omitted from the estimates presented in the present study. Similarly, there were certain spacecraft components such as the RTG on New Horizons that SEER could not estimate and were not included in the estimates. SEER's errors in estimating each of these systems and subsystems are presented, except in cases where the CADRe data did not provide a detailed enough breakdown.

In addition, the average error, weighted average error, median error, and standard deviation of the errors are presented for each WBS and subsystem. The average and median errors are simply the mean and median value of the errors for a given system or subsystem. The weighted average error is the mean value of the errors weighted by the actual costs, which were converted from real-year dollars to a common base year for comparison.

3.1. SEER Point Estimates

SEER is capable of producing both probabilistic cost estimates and point cost estimates. The point estimate is the median value of the probabilistic estimate, meaning there is a 50% chance the estimate will be greater than the actual costs and a 50% chance the estimate will be below the actual costs. This section will examine errors in SEER's point estimates. SEER's errors in estimating mission total, system level, payload total, and spacecraft bus total costs are given in Fig. 3. The missions are ordered by the magnitude of the actual cost for each mission included in this study; where IBEX was the least expensive mission and Juno was the most expensive. Note, that this ordering only takes into account the costs estimated by this study. For example, New Horizons cost more than MESSENGER, but it is ordered between WISE and



	Mission Total	WBS 1, 2, & 3 - PM, SE, SMA	WBS 5 - Payload Total	WBS 6 - Spacecraft Bus Total	Spacecraft PM, SE, IAT	WBS 10 IAT
■ Weighted Average	5%	22%	-1%	-3%	-28%	33%
■ Average	23%	62%	34%	16%	4%	64%
■ Median	-0.3%	12%	31%	2%	-33%	42%
■ Standard Dev	43%	99%	54%	39%	67%	70%
■ IBEX	-16%	-50%	1%	-22%	-62%	128%
■ CONTOUR	99%	264%	64%	92%	166%	165%
■ WISE	64%	95%	84%	-3%	-58%	127%
■ New Horizons	52%	18%		54%	15%	153%
■ MESSENGER	87%	244%	99%			66%
■ GRAIL	-5%	81%	103%			
■ Deep Impact	3%	-37%	32%	6%	-21%	18%
■ MAVEN	25%	77%	30%	20%	-32%	-16%
■ DAWN	-3%	6%		-3%	46%	-19%
■ Kepler	-11%	-19%	-45%	35%	-34%	10%
■ SMAP	-42%	-16%	-47%	-47%	-63%	-1%
■ Juno	-19%	1%	-21%	-23%	-41%	

Figure 3: Percentage error of SEER estimates for mission total, system level, payload total, and spacecraft total costs

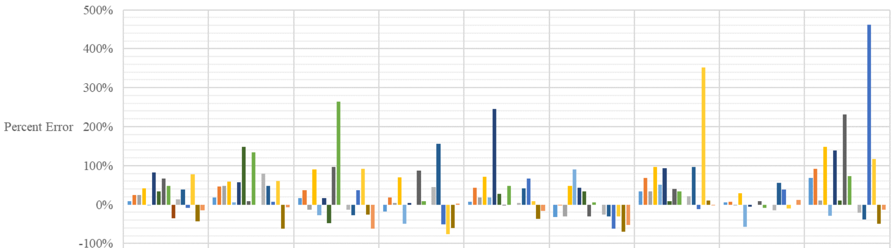
MESSENGER because New Horizons payload, RTG, and operations costs were not included.

The first column of Fig 3 shows SEER's error in estimating total mission cost for each of the twelve missions. Of the twelve missions in the present study, SEER overestimated the cost of six and underestimated the cost of six, resulting in a median error of -0.3%. Therefore, in this aspect SEER was shown to be very accurate because it was just as likely to overpredict the cost as under predict it. The average error was 23%, but the weighted average was only 5%. Thus, SEER was more likely to overestimate the costs of low cost missions, and underestimate the cost of high cost missions. Additionally, when SEER underestimated the cost of a mission it was typically a small error, whereas overestimates tended to be larger errors. For example, the second smallest mission, CONTOUR, was overestimated by 99%, whereas the second largest mission, SMAP, was underestimated by -42%.

Galorath documented that SEER estimates have an average error of -1%, with a standard deviation of 19%. However, it is not known whether the -1% average error from the Galorath study is a weighted average or not. Regardless, when compared to the 5% and 23% weighted average seen in the present study, SEER was not as accurate as in the Galorath study. Additionally, the present study found that the standard deviation of the error in SEER's estimates was 43%, over twice as much as the Galorath study's value of 19%, meaning SEER's estimates were not as precise as in the Galorath study either. However, the Galorath study and the present study were not conducted in the same fashion, with the same goals, and used different data sets. The Galorath study used the data from the end of the mission whereas the present study used predicted masses at CDR. Additional factors that may have contributed to the large errors in the current study vs the results of the Galorath study are discussed in Sec. 4.

The second column of Fig. 3 shows SEER's error in estimating WBS 1, 2, & 3 costs which represent NASAs Project Management (PM), Systems Engineering (SE), and Safety and Mission Assurance (S&MA). The last column shows the error in estimating WBS 10, the final Integration Assembly and Test (IAT) of the spacecraft with its payload and launch vehicle for each mission. All system level costs were estimated using the specific settings recommended by Galorath in their SEER Space Guidance document[7]. However, the weighted average error of WBS 1, 2, & 3 was 22%, and the weighted average error of WBS 10 was 33%. In contrast, the weighted averages for the Payload and Spacecraft Bus were -1% and -3%, respectively. The standard deviation of the WBS 1, 2, & 3 and WBS 10 estimates were 99% and 70%, respectively. In contrast, the standard deviation of the Payload and Spacecraft Bus estimates were 54% and 39%, respectively. Thus, the systems cost estimates at the mission level contributed more error and uncertainty than the payload and spacecraft cost estimates. The fifth column shows the systems level cost for the spacecraft bus. In general, the spacecraft bus's systems costs were underestimated.

Figure 4 shows SEER's errors in estimating each subsystem of the spacecraft bus for each mission. The weighted average error of all the subsystems was 8%. This is in contrast to the spacecraft bus total in Fig. 3 of -3%. Thus, SEER



	Bus Subsystems Total	Structures	Thermal	Propulsion	GN&C	Comm	Electrical Power	Harness	C&DH
Weighted Average	8%	18%	17%	-19%	8%	-31%	35%	6%	68%
Average	24%	46%	38%	18%	43%	-2%	69%	7%	91%
Median	24%	48%	-14%	4%	19%	-30%	34%	-2%	10%
Standard Dev	41%	59%	90%	69%	72%	47%	97%	29%	148%
IBEX	-1%	6%	-27%	-50%	19%	90%	52%	-57%	-29%
CONTOUR	82%	57%	17%	4%	246%	44%	93%	-5%	139%
WISE	34%	148%	-48%		28%	34%	9%		10%
New Horizons	66%	8%	97%	88%	-1%	-31%	40%	9%	232%
MESSENGER	48%	134%	264%	8%	48%	5%	34%	-8%	73%
GRAIL	-36%								
Deep Impact	14%	80%	-14%	45%	4%	-26%	21%	-14%	-22%
MAVEN	39%	48%	-27%	157%	41%	-30%	97%	55%	-38%
DAWN	-9%	7%	37%	-51%	66%	-62%	-12%	38%	462%
Kepler	77%	60%	93%	-75%	9%	-30%	352%	-10%	118%
SMAP	-42%	-62%	-26%	-60%	-37%	-70%	10%	0%	-48%
Juno	-15%	-6%	-61%	3%	-16%	-52%	0%	12%	-13%

Figure 4: SEER Subsystems Comparison

underestimated PM, SE, and IAT costs associated with a contractor building a spacecraft bus, but overestimated the costs of designing and producing the subsystem components. SEER overestimated the cost of the structures and mechanisms for all missions except SMAP and Juno, the two largest missions. SEER underestimated the cost of half of the thermal subsystems, but when it overestimated the costs, the magnitude of the error was larger, giving the thermal subsystems an overall weighted average error of 17%. The propulsion subsystems were typically overestimated, but when they were underestimated the error was less than -50%. This brought the weighted average of the propulsion subsystem error down to -19%. GN&C errors were relatively small with the exception of CONTOUR. However, the GN&C subsystem of CONTOUR was only 4% of the subsystems' cost, so while the percent error was very large the absolute error was only moderate. SEER underestimated the majority of the communication subsystems costs. The communication subsystem is the subsystem most underestimated by SEER with a weighted average error of -31%. Electrical power was consistently overestimated except in the case of Dawn, the only mission in this study which utilized solar electric propulsion. The electrical power subsystem was the second most overestimated subsystem by SEER, with a weighted average error of 35%. The harness subsystem cost was most accurately estimated by SEER. The harness subsystem's weighted average error, median error and standard deviation are all smaller in magnitude than all other subsystems. The C&DH subsystem, in contrast, had the largest error and standard deviation of all the subsystems. The large error of the C&DH subsystem is likely due to the large uncertainties in estimating custom integrated circuits, such as Field Programmable Gate Arrays (FPGAs).

3.2. SEER Probabilistic Estimates

This section examines SEER's probabilistic cost estimates. For every input in SEER the user inputs a "least," "likely," and "most" value corresponding to an optimistic, most likely, and pessimistic assumption about the input. SEER models uncertainty by assigning each "work element" a distribution of possible costs in addition to a mean cost. The least/likely/most inputs for each work element correspond to the lower bound, mode, and upper bound of the Probability Density Function (PDF) of a beta distribution given by

$$f(x) = \begin{cases} \frac{x^{(\alpha-1)}(1-x)^{(\beta-1)}}{B(\alpha,\beta)} & 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where $B(\alpha, \beta)$ is the beta function which itself is a function of gamma functions given by

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)} \tag{3}$$

α and β are shape parameters given by

$$\alpha = \frac{(\mu - i_{least})(2i_{likley} - i_{least} - i_{\mathrm{most}})}{(i_{likely} - \mu)(i_{most} - i_{least})} \tag{4}$$

and

$$\beta = \alpha\frac{(i_{most} - \mu)}{(\mu - i_{least})}. \tag{5}$$

Where i_{least}, i_{likely}, and i_{most} are the least/likely/most inputs into SEER, and μ is the mean of the distribution given by

$$\mu = \frac{i_{least} + 4i_{likely} + i_{most}}{6} \tag{6}$$

When SEER runs a Monte Carlo simulation it selects the value of the inputs to its CERs in a random way determined by Eq. 2. The results of the Monte Carlo simulation are then used to create a cumulative distribution function (CDF) for the cost of the mission.

The frequency that the actual cost of a mission was within SEER's 80% confidence interval was used to evaluate the accuracy of SEER's uncertainty quantification capabilities. A confidence interval is the interval between two points, or confidence levels, on a CDF. When a model claims an estimate has an 80% confidence interval, that means 80% of the time the actual value should be within that interval. Thus, based on SEER's uncertainty quantification output, it is expected that actual costs of nine or ten of the twelve missions (approximately 80%) will be within SEER's 80% confidence interval. Because SEER only outputs the values of the CDF, known as confidence levels, at intervals of 10%, the user must be content with only knowing the 80% confidence interval instead of the more common 90% or 95% confidence intervals. The 80% confidence intervals generated by SEER for the mission total cost, payload total cost, and spacecraft bus total cost can be seen in Figs. 5, 6, and 7. Figure 5 shows that the only missions with actual costs outside SEER's 80% confidence interval were CONTOUR, WISE, and MESSENGER. Since twelve missions were included it would be expected that two or three would fall outside of the confidence interval which is exactly what was observed. It can be concluded that while the point estimates given by SEER had a wide variance, the CI's generated by SEER for the total mission cost captured the actual cost the expected number of times.

It should be noted that the confidence intervals seen in Fig. 5 are extremely large. This was primarily due to the uncertainty in modeling the spacecraft from the data available. In particular, it was extremely difficult to judge the level of heritage of a majority of the spacecraft components. In some cases,

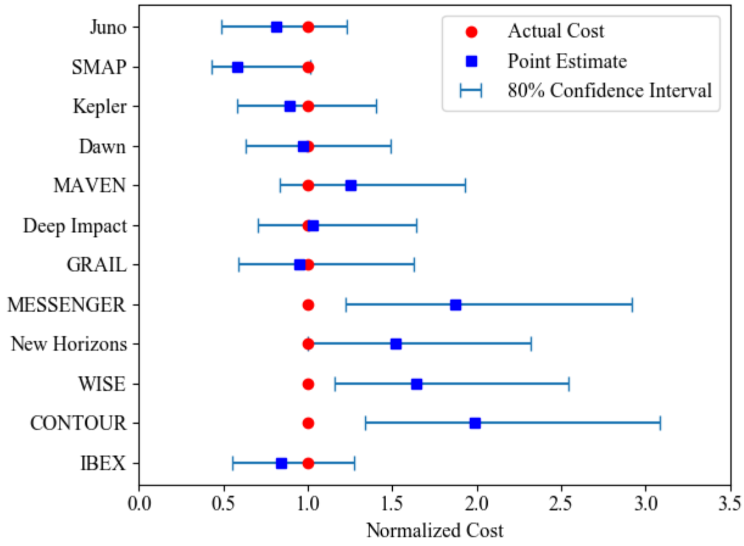


Figure 5: SEER 80% confidence intervals for total mission costs.

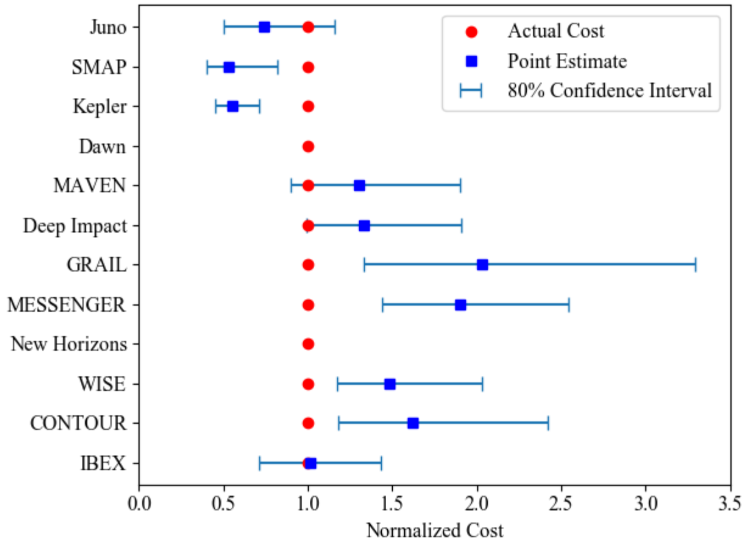


Figure 6: SEER 80% confidence intervals for total payload costs. The payloads of Dawn and New Horizons were not modeled.

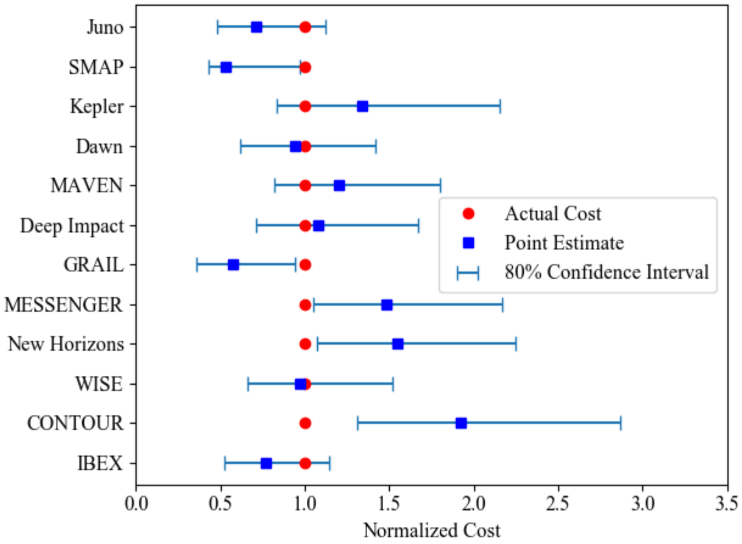


Figure 7: SEER 80% confidence intervals for total spacecraft bus costs.

there was no information on whether a component was a brand new design, or a copy of a previously flown component. In these cases, the estimators would adjust the new design least, likely, and most inputs to reflect the uncertainty.

The 80% confidence intervals generated by SEER for the payload total cost can be seen in Fig. 6. As previously stated, because there was not enough information available to produce a cost estimate of New Horizons and Dawn's payload costs, those estimates are omitted from Fig. 6. With the ten remaining payloads it would be expected that eight of the ten would fall within the confidence interval, however, only four of the payloads do.

The 80% confidence intervals generated by SEER for the spacecraft bus total cost can be seen in Fig. 7. Since all twelve spacecraft are included it is expected that 2-3 will fall outside of the 80% confidence interval. However, five spacecraft fall outside of the confidence interval. These results will be discussed in the following section.

4. Discussion

In this section, a number of factors which may have adversely affected the outputs of SEER in this blind study are discussed. Results presented in this section were produced after the results in the previous section (Sec. 3) were analyzed and therefore are not blind. Some of what is discussed in this section is the interpretation of the results by the authors based on their experience as cost estimators and has not been rigorously tested.

4.1. Mass Estimation and Related Uncertainty

In SEER, mass is the primary scaling mechanism for the majority of components in the structures, thermal, and propulsion subsystems, as well as a few components in other subsystems. For the present study, it was assumed that the low mass input was the current best estimate, the likely estimate was the current best estimate plus the mass contingency, and the high estimate was 30% above the likely estimate. These assumed mass inputs are common in the cost estimating community and are recommended by the SEER Space Guidance document[7]. Despite being a common assumption in the cost estimating community, it is not clear where the assumption that the high mass estimate should be equal to the likely assumption plus 30% originates. However, this assumption is consistent with the discussion in paper by Wilhite et al. which showed that the average spacecraft increases in mass by 28.5% from its preliminary design to launch[20]. The 30% assumption does not take into account the design maturity of the components, or how close the mission is to completion. The Jet Propulsion Laboratory mass margins approach defined in their handbook of design principles for flight systems bases the mass margin on what phase the mission is in. The handbook recommends mass margins of 30% at PMSR (Preliminary Mission & Systems Review), and 10% at CDR[21]. The high 30% mass margin used for the present study is likely a contributing factor, for the high average error seen in the SEER point estimates. However, the high 30% mass margin also assures that the confidence intervals are large enough to capture the actual mission costs.

To test if SEER would perform better with the JPL recommended mass margins a Python script was written to generate SEER command files to alter the mass inputs of the structural/mechanical components for each spacecraft/payload. Figure. 8 shows the median estimates and confidence intervals with a mass margin of 10% instead of 30%. While the point estimates improved marginally from previous results the confidence intervals narrow and upper bound of SMAP's confidence interval slips below the actual cost of the mission. Since lowering the upper mass margin does not significantly effect the lower bound of the estimates the lower bounds of CONTOUR, WISE and MESSENGER do not slip below their actual costs either. This results in there now being only eight missions (67%) whose actual costs fell in the 80% confidence interval of the SEER estimates.

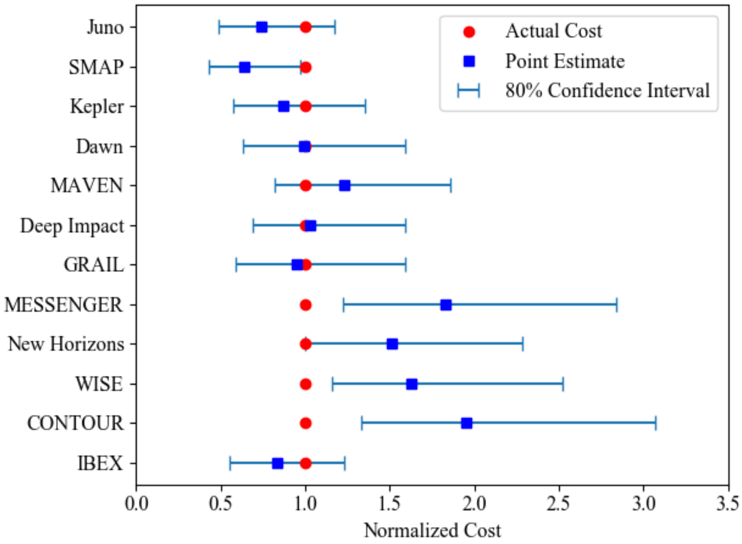
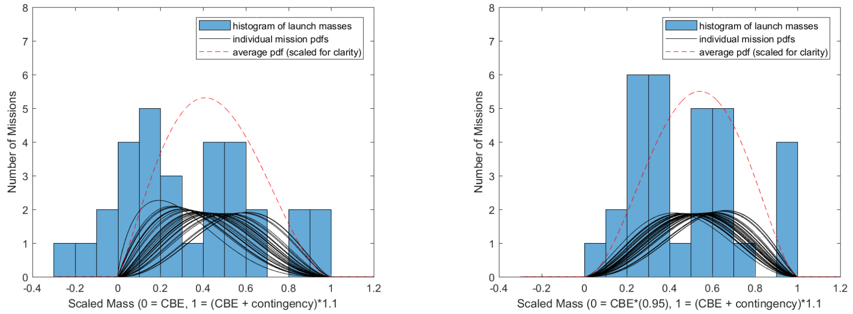


Figure 8: SEER 80% confidence intervals for total mission costs with 10% mass margin.

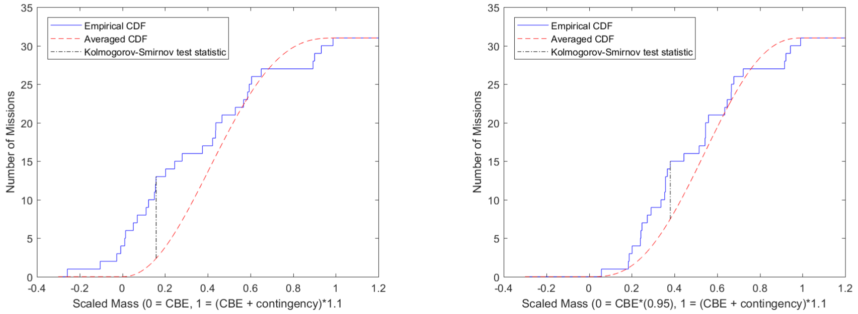
Returning to Fig. 5, three of the twelve missions were overestimated by SEER to the point that lower bounds of the confidence interval exceeded the actual costs of the mission. Thus, it is likely that the lower bound of the mass estimate is not low enough, rather than the upper bound of the mass estimate being too high. Recall that it is common practice and explicitly recommended by Galorath to use the Current Best Estimate (CBE) of the mass to define the lower bound of a beta distribution, the CBE plus contingency mass to define the mode, and a mass margin added to the mode to define the upper bound. By defining the CBE as the lower bound this assumes there is a 0% chance that the spacecraft mass will decrease. However, Fig. 9a shows this is not true.

Figure 9a compares the beta distributions defined by these mass estimates with a histogram of the launch masses for 31 separate NASA missions. The x-axis is scaled so that the CBE of the mass of each mission at CDR is equal to zero and the CBE plus mass contingency multiplied by the mass margin is equal to one, while the y-axis represents the number of missions. The 31 beta distributions representing the predicted mass of the 31 NASA spacecraft are shown as black lines and the average of the 31 distributions is shown as a red dashed line. The average distribution has been multiplied by a factor of three to make it clearly visible on the figure. While the majority of mission's launch masses were within the predicted range, four of the launch masses were lower than the CDR current best estimate. SAGE III was the mission whose launch mass was furthest below CBE, launching at 96% of CBE. Given that the conventional mass modeling assumptions assign a 0% probability to the launch mass being less than the CBE at CDR, the mass modeling assumptions are invalid and should be modified. Figure. 9b shows a new set of beta distributions using mass modeling assumptions which have been modified from Fig. 9a such that the lower bound of the distribution is 95% of the CBE rather than CBE. Figures 9c and 9d compare the empirical CDF defined by the launch masses with the averaged CDF defined by the mass modeling assumptions. Figure 9c compares the empirical CDF of the launch masses with the averaged CDF defined by the conventional mass modeling assumptions while Fig. 9d compare the empirical CDF with the averaged CDF defined by the modified mass modeling assumptions. The y-axis is scaled to the number of missions as opposed to being scaled from zero to one as is common practice with CDFs. Both Figs. 9c and 9d display the Kolmogorav-Smirnov test statistic, which is a method of determin-



(a) Conventional mass modeling assumptions at CDR.

(b) Modified mass modeling assumptions at CDR.



(c) Conventional mass modeling assumptions at CDR.

(d) Modified mass modeling assumptions at CDR.

Figure 9: Comparison of beta PDFs and CDFs used to predict mission mass at CDR with actual mission mass.

ing the goodness of fit of an empirical CDF to a hypothesized CDF and is simply the maximum difference between the CDFs[22]. From visual inspection of Figs. 9c and 9d the CDF using the modified mass modeling assumptions fits the data better than the CDF using the conventional mass modeling assumptions. Furthermore, the Kolmogorav-Smirnov test statistic for the conventional assumptions is 10.6 whereas it is only 7.5 for the modified assumptions.

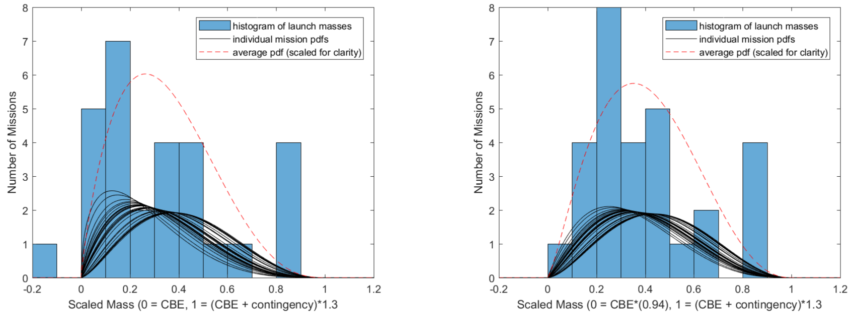
The Minkowski L_1 metric or area metric is more robust than the Kolmogorav-Smirnov test statistic and is given by

$$d(F, E_n) = \int_{-\infty}^{\infty} |F(x) - E_n(x)| dx \tag{7}$$

where $F(x)$ is the predicted distribution and $E_n(x)$ is the empirical distribution[23]. The area metric is commonly used to validate statistical methods and measures the area between the empirical and hypothesized CDF. The area metric for the conventional CDF was 4.4 and was 2.3 for the modified CDF. Again the modified mass modeling assumptions outperformed the conventional ones.

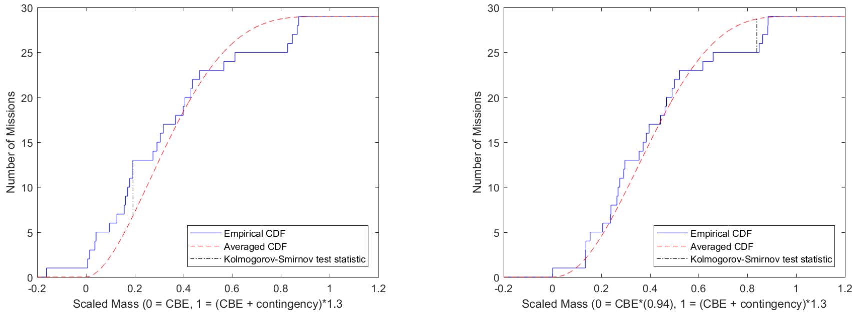
Figure 10 shows the same types of plots and tests as Fig. 9 but instead compares predictions at PDR with launch masses for 29 NASA missions. The conventional mass modeling assumptions at PDR are the same as at CDR only with a 30% mass margin defining the upper bound instead of 10%. Figure 10a shows that these bounds do not encompass all the missions. SAGE III had a launch mass 6% less than the CBE at PDR. Thus, the lower bound is adjusted to be 94% CBE as seen in Fig. 10b. Figure 10c show the empirical CDF defined by the actual launch masses compared to the averaged CDF defined by the conventional mass estimating assumptions which has a Kolmogorav-Smirnov test statistic of 6.2 and an area metric of 2.3. Figure 10d shows the same emprical CDF but compared to the averaged CDF defined by the modified mass estimating assumptions which now has a Kolmogorav-Smirnov test statistic of 3.8 and an area metric of 1.3.

While it is uncommon for missions to launch with a lower mass than the CBE at PDR or CDR it is does happen and thus it is invalid for cost estimators exclude the possibility. The authors recommend that the conventional mass



(a) Conventional mass modeling assumptions at PDR.

(b) Modified mass modeling assumptions at PDR.



(c) Conventional mass modeling assumptions at PDR.

(d) Modified mass modeling assumptions at PDR.

Figure 10: Comparison of beta PDFs and CDFs used to predict mission mass at PDR with actual mission mass.

modeling assumptions be modified such that at PDR the lower bound is 94% of CBE and at CDR the lower bound is 95% of CBE. The SEER estimates for the twelve missions in the blind study were recomputed such that the mass estimates for all mechanical/structural components had a lower bound of 95% of the CBE instead of the CBE. This lowered the point estimates and lower bounds of the estimates slightly but no additional missions fell within the confidence intervals of the estimates. While mass is primary input used to scale the cost of mechanical components it does not always have as big of an effect on cost as other inputs such as percent new design. However, depending on the details provided in the documentation determining these other inputs can be challenging without the help of subject matter experts as discussed in the following subsection.

4.2. Subject Matter Experts

The role of subject matter experts is critical in any estimation. Unfortunately, since the missions in the present study took place over the past 20 years it was not possible to contact any of the subject matter experts who worked on these missions to clarify details missing in the CADRes or supporting documents. As a result, there was a wide range of uncertainty in the inputs and results. One of the subsystems driving this overestimation was the C&DH subsystem. Each of these spacecraft shared a common C&DH electronics module with significant heritage. The documents from CDR described the electronics thoroughly, including circuit diagrams, and technical specifications. However, they did not mention that the majority of the circuit boards were copied from previous missions. During the blind estimation, it was assumed that these electronics boards were modified to fit the requirements of their respective missions, however, it is more likely that the electronics were copies of previous flight units with only one board being unique to each mission. A five minute conversation with an electrical engineer from any of these missions could have cleared up the confusion and significantly improved the accuracy of the modeling.

4.3. Prototyping Assumptions

The number of assumed prototypes can have a significant effect on the cost estimate. In SEER, each additional prototype adds to the development cost. A standard assumption in the cost estimating community is a prototype input value of 1.3. A fractional prototype represents a prototype which does not have the full functionality of the working component, such as a breadboard, engineering design unit, or boiler plate unit. For many spacecraft components there were no details provided on how many prototypes were made during mission development. When there was no data on the number of prototypes, the assumed value was 1.3. It is unclear where the assumption that the average spacecraft component for a NASA science mission will have 1.3 prototypes originated. The authors plan to investigate the effect of the number of prototypes on the accuracy of cost estimates in the future. The standard assumption of 1.3 prototypes may have inflated the SEER estimates as it is possible that in some cases the reason there was no prototype information was documented for certain components was because there were no prototypes for those components.

4.4. Experience of Estimators

The experience and skill of the estimator also plays a significant role in the accuracy of a cost estimate. Estimators B and C were trained to use SEER and completed a number of training exercises for several weeks before the study. In addition, estimators B and C were advised by more experienced estimators who were on hand to answer any questions they might have. However, we would be remiss not to mention that a cost estimator with several years of experience would likely have produced more accurate and precise results with SEER.

5. Conclusions

The present work evaluated the parametric cost estimating tool SEER-H and its capability to model the cost of NASA space missions through a blind study. The study was blind in that the estimators had no knowledge of the actual costs of the missions being estimated. This was done to prevent cognitive biases, such as anchoring, from influencing the way the SEER was used to estimate the costs.

The present study found that SEER had an average error of 23%, median error of -0.3%, and a standard deviation of 43%. Weighing the errors by the actual cost of the mission the average error was only 5%. The present study also determined that SEER tended to overestimate smaller missions and underestimate larger missions. SEER's uncertainty quantification capabilities worked well at the mission level; of the twelve missions in the present study, nine of the missions' actual costs were within the 80% confidence interval given by SEER. However, at the payload and spacecraft levels only 40% and 58% of the actual costs fell with in the 80% confidence interval of SEER. One major observation was that in general the larger the mission and the more items estimated the more accurate SEER's estimates would be.

There are several factors which may have affected the results of this study. These factors include the assumptions about mass margins, numbers of prototypes, as well as the level of detail provided in the documentation of the missions and the experience of the estimators. The effects of mass margin assumptions were explored in detail. It was determined from comparing actual masses of 31 spacecraft with the predicted mass earlier in the design process that the current mass margin assumptions should be modified to include the possibility that the spacecraft may decrease in mass before launch as this happens in some cases. The authors recommend that in the future, the lower bound of the mass estimate be 94% of CBE at PDR and 95% of CBE at CDR as rather than the typical assumption of the lower bound being the CBE.

Another factor which likely had a large effect on the results of the present study was that the estimators were not able to ask clarifying questions from subject matter experts involved in the missions. The estimators had to rely on documents presented at major reviews only, which often omitted important information; such as, the heritage of a particular component.

References

[1] J. R. Wertz, W. J. Larson, Reducing space mission cost, Microcosm Press Torrance, CA, 1996.

[2] H. Apgar, Cost estimating, Space Mission Engineering: The new SMAD.

[3] S. J. Kapurch, NASA systems engineering handbook, Diane Publishing, 2010.

[4] NASA, 2008 cost estimating handbook, Washington, DC.

[5] O. Trivailo, M. Sippel, Y. Şekercioğlu, Review of hardware cost estimation methods, models and tools applied to early phases of space mission planning, Progress in Aerospace Sciences 53 (2012) 1–17.

[6] S. Sanchez, K. Kha, SEER Validation Study Results for NASA Space Science Missions (2015).

[7] Galorath, SEER-H Space Guidance: Revision 2.2 (2016).

[8] D. McComas, F. Allegrini, P. Bochsler, M. Bzowski, M. Collier, H. Fahr, H. Fichtner, P. Frisch, H. Funsten, S. Fuselier, et al., IBEX–interstellar boundary explorer, Space science reviews 146 (1-4) (2009) 11–33.

[9] J. Veverka, R. Farquhar, E. Reynolds, M. Belton, A. Cheng, B. Clark, J. Kissel, M. Malin, H. Niemann, P. Thomas, et al., Comet nucleus tour, Acta Astronautica 35 (1995) 181–191.

[10] T. Bradley, et al., Contour mishap investigation board report.

[11] F. Liu, R. Cutri, G. Greanias, V. Duval, P. Eisenhardt, J. Elwell, I. Heinrichsen, J. Howard, W. Irace, A. Mainzer, et al., Development of the wide-field infrared survey explorer (WISE) mission, in: Modeling, Systems Engineering, and Project Management for Astronomy III, Vol. 7017, International Society for Optics and Photonics, 2008, p. 70170M.

[12] J. C. Leary, R. F. Conde, G. Dakermanji, C. S. Engelbrecht, C. J. Ercol, K. B. Fielhauer, D. G. Grant, T. J. Hartka, T. A. Hill, S. E. Jaskulek, et al., The messenger spacecraft, in: The MESSENGER Mission to Mercury, Springer, 2007, pp. 187–217.

[13] M. T. Zuber, D. E. Smith, D. H. Lehman, T. L. Hoffman, S. W. Asmar, M. M. Watkins, Gravity recovery and interior laboratory (grail): Mapping the lunar interior from crust to core, in: GRAIL: Mapping the Moons Interior, Springer, 2012, pp. 3–24.

[14] W. H. Blume, Deep impact mission design, Space Science Reviews 117 (1-2) (2005) 23–42.

[15] B. M. Jakosky, R. Lin, J. Grebowsky, J. Luhmann, D. Mitchell, G. Beutelschies, T. Priser, M. Acuna, L. Andersson, D. Baird, et al., The Mars atmosphere and volatile evolution (MAVEN) mission, Space Science Reviews 195 (1-4) (2015) 3–48.

[16] V. C. Thomas, J. M. Makowski, G. M. Brown, J. F. McCarthy, D. Bruno, J. C. Cardoso, W. M. Chiville, T. F. Meyer, K. E. Nelson, B. E. Pavri, et al., The dawn spacecraft, in: The Dawn Mission to Minor Planets 4 Vesta and 1 Ceres, Springer, 2011, pp. 175–249.

[17] G. Basri, W. J. Borucki, D. Koch, The kepler mission: A wide-field transit search for terrestrial planets, New Astronomy Reviews 49 (7-9) (2005) 478–485.

[18] D. Entekhabi, E. G. Njoku, P. E. O'Neill, K. H. Kellogg, W. T. Crow, W. N. Edelstein, J. K. Entin, S. D. Goodman, T. J. Jackson, J. Johnson, et al., The soil moisture active passive (SMAP) mission, Proceedings of the IEEE 98 (5) (2010) 704–716.

[19] S. Matousek, The juno new frontiers mission, Acta Astronautica 61 (10) (2007) 932–939.

[20] A. Wilhite, D. Reeves, D. Stanley, J. Wagner, Evaluating the impacts of mass uncertainty on future exploration architectures, in: Space 2006, 2006, p. 7250.

[21] N. Yarnell, Design, verification/validation and operations principles for flight systems, Jet Propulsion Laboratory, D-17868, Rev 2.

[22] L. J. Bain, M. Engelhardt, Introduction to probability and mathematical statistics, Brooks/Cole, 1987.

[23] W. L. Oberkampf, C. J. Roy, Verification and validation in scientific computing, Cambridge University Press, 2010.