Geo-Statistical Approach to Estimating Asteroid Exploration Parameters

William Lincoln, Jeffrey H. Smith, Charles Weisbin

Jet Propulsion Laboratory, California Institute of Technology

4800 Oak Grove Drive, Pasadena, CA 91109

William.Lincoln@jpl.nasa.gov

Abstract

NASA's vision for space exploration calls for a human visit to a near earth asteroid (NEA). Potential human operations at an asteroid include exploring a number of sites and analyzing and collecting multiple surface samples at each site. In this paper two approaches to formulation and scheduling of human exploration activities are compared given uncertain information regarding the asteroid prior to visit. In the first approach a probability model was applied to determine best estimates of mission duration and exploration activities consistent with exploration goals and existing prior data about the expected aggregate terrain information. These estimates were compared to a second approach or baseline plan where activities were constrained to fit within an assumed mission duration. The results compare the number of sites visited, number of samples analyzed per site, and the probability of achieving mission goals related to surface characterization for both cases.

Introduction

To investigate the processes that marked the initial stages of planet and satellite formation, a potential precursor mission would formulate hypotheses concerning the deposition of minerals over different surface types. Subsequent surface operations during a human visit would generally result in acceptance or rejection of those hypotheses depending on the number of samples analyzed per site, the number of sites visited, and the required confidence level associated with the decision. The ability to locate "unexpected" samples (i.e. those opposed to the precursor hypotheses) was used as the measure of mission science success in this paper. The analysis output was a recommended plan given all existing information. The plan was efficient, in the sense of allocating the appropriate amount of time at each site, and visiting the fewest number of sites to reach the mission goals to within a given confidence level.

Before a precursor asteroid mission, available information about asteroids with known maps can be used as analogs to develop planning scenarios. After a precursor asteroid mission, precursor maps would then provide a basis for feasible human exploration scenarios. There are, in general, three key aspects to be drawn from precursor data: (1) enable safe human operations; (2) mission assurance; and (3) scientific data. The focus of this paper was on scientific hypotheses developed from all precursor data. Clearly, the subsequent human operations would be driven in large part, by activities designed to assess the validity of the precursor hypotheses.

The approach (Weisbin et al. 2008) developed in this study began with fundamental science questions (e.g. deducing initial stages of formation through examination of composition) (Fellows, 2007). Next, using the expected precursor derived information (e.g. size, porosity, chemical composition), the number of observed homogenous regions on the asteroid were estimated from precursor maps. With this geological characterization, the number of sites to be visited and the number of samples per site were estimated in order to find, with a specified probability, a sample opposed to the precursor hypotheses. A sample opposed to the precursor hypotheses is particularly interesting since it represents the discovery of a sample previously thought unlikely based on all the precursor information. From the number of sites and samples, the mission duration and experimental equipment needed for a human mission is derived.

Background

Fundamental Science Questions. Visiting an asteroid would address a number of scientific questions, and in particular the sampling strategy for an asteroid could address the following questions (Fellows, 2007):

- What processes marked the initial stages of planet and satellite formation (bulk composition versus solar distance, interior structure and evolution)?
- How did impactor flux decay during the solar system's youth and influence the timing of life's emergence on Earth? What was the history and role of early impacts, impactor flux in the early solar system, and calibration of impact record? How impacts altered the asteroid's history, evolution, and orbital dynamics?
- What is the history of volatiles, especially water (distribution, character, origin of volatiles, potential resources)?
- How do the processes that shape the contemporary character of planetary asteroids operate and interact (absolute ages of samples, recent cratering history and current flux, potential resources)?

Existing Asteroid Maps. The approach presented here would normally use actual asteroid maps if available. Currently when detailed maps are not yet available for a particular destination, asteroids with known maps are used as proxy's to develop scenarios. While much is unknown about asteroids in general and the particular asteroids that will be visited by human explorers, unmanned missions have remotely sensed some asteroids. A number of asteroids have been imaged by spacecraft: Gaspra (Galileo), Ida (Galileo), Eros (NEAR), Mathhilde (NEAR), Itokawa (Hayabusa), Lutetia (Rosetta). The validity of using existing maps depends on the similarity (i.e. body size, distribution of regions) between the asteroids for which information is available and the asteroid mission candidate. When detailed maps for the asteroid of interest become available from precursor missions then those maps can be used to refine the scenarios. Today without detailed maps of potential candidate asteroids of interest, the number of sites and samples were estimated by extrapolation from maps of asteroid analogs. Consequently, the results assumed a range of geo-statistical distributions (Britt, D.T. et al. 2006).

The problem addressed in this paper was how to obtain an estimate of the number of sites and samples to explore without a detailed map? With no detailed maps of potential asteroids, a geo-statistical analysis approach was developed. A range of geo-statistical distributions of asteroids was assumed and although the approach was limited by the model fidelity and validity to some extent, it allowed a pathway for extrapolation to other asteroids.

Itokawa Regions. The Hayabusa mission to asteroid Itokawa was used as a surrogate for a priori maps. Eros has also been mapped but not to the same resolution as Itokawa (Thomas et al. 2002). A number of hypotheses have been postulated based on Hayabusa data. For example data of Itokawa suggested representation of 34 regions: 1 large boulder, 7 craters, 3 regolith regions, 22 rough/boulder fields, and in addition there is a location with unusual albedo. Hayabusa found no substantial difference in mineralogical composition over the whole surface. It has been hypothesised that chemical composition within a region class is homogeneous (Saito et al. 2006), (Demura et al. 2006). The data suggests that the surface is likely comprised of homogeneous material (likely LL- or L-chondrites) (Okada, 2006).

Target number of samples per site to be used subsequently for comparison to our results. Potential surface operations timelines associated with various design reference missions (DRM) have been developed for generic asteroid candidates (HEFT Phase II. 2010). For example, a DRM with 30 days to spend at an asteroid has (in addition to surface analysis and sample collection) a number of other proposed surface activities. These activities include active seismometry, drilling and retrieving drill core, deploying science packages and radar sounder, ISRU demonstration, and planetary defense experiments. This analysis focused only on the time allocated for surface analysis and sample collection. Assuming 2 minutes for the astronaut to examine a sample (Heiken G. et al. 2007), Table 1 gives upper bounds on the number of samples per site based on the DRM timeline of surface operations.

Table 1. Upper bounds on the number of samples analyzed within a site assuming a 2-minute per sample analysis time.

Site	0	1	2	3	4	5
Number of samples per site	40	50	80	100	190	290

Approach

Each region class has an assumed number of candidate sites of which a subset would be visited. Each region class, e.g. rough/boulder fields, has similar geographic characteristics (Demura, 2006). We assume a site is a small area (e.g. ~10m radius accessible by an anchored astronaut) within a region. The Hayabusa mission identified a number of regions within each region class and hypothesized they were likely to have similar mineral compositions. To confirm/reject this homogeneity hypothesis a number of candidate sites spread over regions of the same type was assumed. For a given required confidence level, the number of sites to visit within a region class was estimated in order to locate a sample of different chemical composition (if one existed) based on the probability of finding such a sample at each site visited.

One of the strengths of human exploration is the ability of astronauts to discover scientifically "interesting" samples. Here we define interesting to be a sample opposed to the hypotheses developed based on the precursor data. Let a type 0 sample denote a sample consistent with the precursor hypothesis, and a type 1 sample is opposed to the precursor hypothesis.

Figure 1 shows an abstract representation of the model's assumptions, and illustrates its parameters.

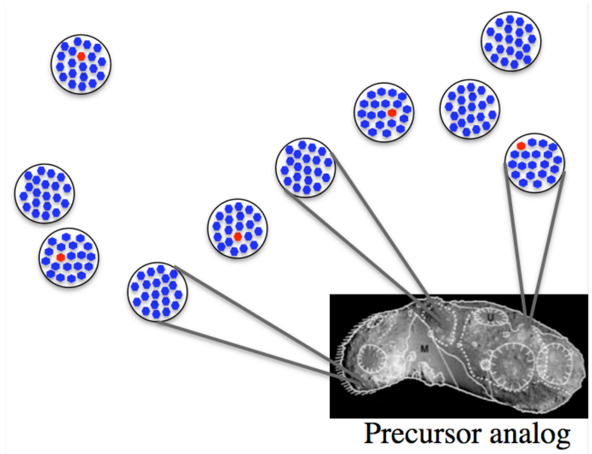


Figure 1. Illustration of the model assumptions. A circle containing a number of potential samples represents each candidate site. The blue samples represent type 0 samples, and the red samples represent the type 1 samples. Note that not all sites are assumed to have type 1 samples. The arrangement of the sites relative to each other is illustrative and not an input to the model in this paper. Precursor analog is Hayabusa mission to Itokawa (Demura 2006).

The minimum number of samples is determined by selecting the number for which, if such a type 1 sample existed, it would likely be found given a priori estimates of the likelihood of being there.

Model Input Parameters. The input parameters were:

1. *nRegions*: Number of region classes on asteroid.
2. *nCandidateSites*: Number of candidate sites over each region class.
3. *q*: Likely fraction of the candidate sites with some type 1 samples.
4. *nPotentialSamples*: Number potential samples per site.
5. *p*: Likely fraction of samples at a site that are type 1, given there are type 1 samples at that site.
6. π: Required probability threshold of finding at least one type 1 sample.

This kind of data comes from precursors such as Hayabusa to Itokawa.

Model Outputs. The model computes for each possible value of the number of visited sites (*nVisited*), the minimum number of samples to analyze per site (*nAnalyzed*) to exceed the required probability, π, of finding at least one type 1 sample during the human mission.

Steps of the Implemented Model.

1. For each region class = 1,..,*nRegions*:
    2. For each possible number of visited sites, *nVisited*=1,..,*nCandidateSites*:
        3. For each possible number of samples analyzed per site=1,..,*nPotentialSamples*:
            4. Compute probability of finding at least one type 1 sample for the parameter combination: (*nVisited*, *nAnalyzed*).
            5. Output set of {(*nVisited*, *nAnalyzed*)} that meet the required probability threshold, π.
    6. The recommended total number of sites during the mission is computed by summing over all the different region classes.

The calculation of the probability of finding at least one type 1 sample (step 4) is developed in the next two sections.

Number of Samples Examined per Site. Our model assumes sampling without replacement within a site since samples, once taken, would not be resampled. The parameter *nPotentialSamples* is finite considering the assumed ~10m surface site. Later, the results will show the output is insensitive to this parameter over the values considered in this study.

The probability of finding a type 1 sample at a single site as a function of the actual number of samples examined per site was estimated parametrically using a hypergeometric probability model (Rohatgi 2003). Then the minimum number of samples per site needed to achieve the required probability threshold, π, of finding at least one type 1 sample was computed.

To illustrate this calculation suppose p = fraction of samples not being LL5 (a chondrite type (Van Schmus, W. R. et al. 1967)) when the precursor hypotheses suggests samples within the site are LL5. If p=0, all samples in the site are type 0. For p>0, there are *p*nPotentialSamples* type 1 samples and *(1-p)*nPotentialSamples* type 0 samples. Then the probability of finding at least one such type 1 ("unexpected") sample at a single site is computed from the hypergeometric distribution (Rohatgi 2003).

pSite(as a function of *nAnalyzed*)

= probability of finding at least one type 1 sample at a given site (1)

= 1 – probability finding no type 1 samples = $1-\frac{\binom{(1-p)*nPotentialSamples}{nAnalyzed}}{\binom{nPotentialSamples}{nAnalyzed}}$

Where the parameter *nAnalyzed* = number of samples selected and analyzed per site, $(1-p)*nPotentialSamples \geq nAnalyzed > 0$, and $\binom{N}{r}$ is the number of r-combinations of a set with N elements and equals N!/r!(N-r)!. Note if $nAnalyzed > (1-p)*nPotentialSamples$ then it can be shown that at least one type 1 sample would be found.

Number of Sites Visited per Region Class. This section describes the calculation of the probability of finding at least one type 1 sample in all the visited sites of the same region class.

Let *A* be the event "Do not find any type 1 samples, from a given region class, during the mission". Let H_j be the event "Visit exactly j sites, from a given region class, with some type 1 samples". Then conditioning on the events $\{H_j\}$:

$$P(A) = \sum_{j=0}^{number\ of\ visited\ sites} P(A \mid H_j)P(H_j) \quad (2)$$

$P(A \mid H_j)$ is the probability of finding at least one type 1 sample, at a given site, given that exactly j sites are visited. Using equation (1), $P(A|H_j) = (1\text{-}pSite)^{j} * 1^{(nVisited - j)}$

= $(1\text{-pSite})^{j}$ (since exactly j sites have type 1 samples, there is probability 1 of not finding any type 1 samples at the other, $nVisited - j$, sites. Analogous to sampling without replacement for samples within a site, the visited sites are sampled without replacement (Rohatgi, 2003) from the candidate sites of a given region class:

$$P(H_j) = \frac{\binom{q*nCandidateSites}{j}\binom{(1-q)*nCandidateSites}{nVisited-j}}{\binom{nCandidateSites}{nVisited}}$$

where *q*nCandidateSites* is just the number of candidate sites with some type 1 samples, and *(1-q)*nCandidateSites* is the number of candidate sites with no type 1 samples. Then probability of finding at least one type 1 sample, in the particular region class, during the mission, is equal to 1-P(A).

This calculation yields the smallest number of samples to analyze per site (*nAnalyzed*) for each value of the number of visited sites (*nVisited*) that gives at least the required probability, π, of finding at least one type 1 sample.

Results

Number of Samples per Site. Figure 2 plots the probability of finding at least one type 1 sample at a given site, pSite, versus the number of actual samples analyzed per site, *nAnalyzed*. The input parameter *nPotentialSamples*=1000. From Figure 2, the trade-off between the number of samples analyzed and chance of finding an unexpected sample can be viewed. For example, with *p*=0.05, when the number of samples analyzed per site equals 44, the probability of finding a least one type 1 sample at a given site is 90%. For smaller values of *p* the number of analyzed samples is higher. For example, with *p*=0.01, 205 samples are needed to ensure at least 90% chance of finding at least one type 1 sample at the site.

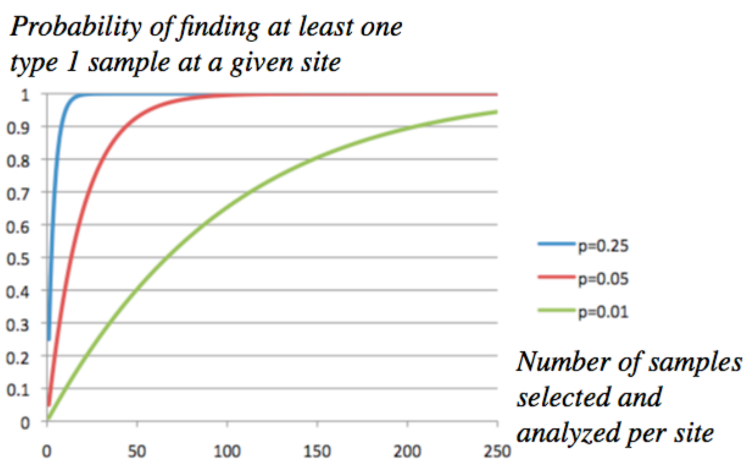
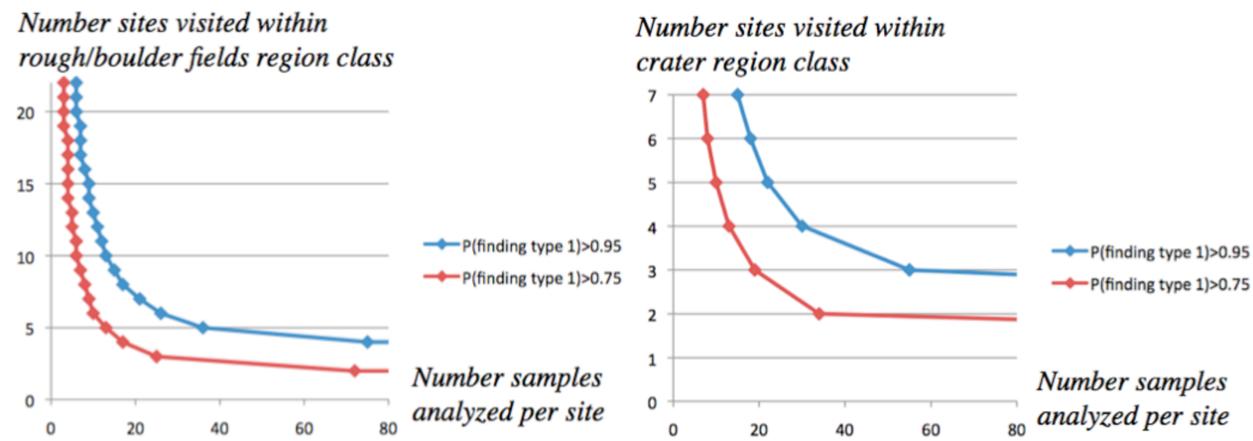


Figure 2. Probability of finding a type 1 sample within a site, *pSite*, as function of the number of samples analyzed at site, *nAnalyzed*, for p=0.01, 0.05, and 0.25.

Number of Sites per Region Class. Figure 3 shows the number of sites visited, within specific region class (rough/boulder field), versus the number of samples analyzed to achieve a 75% and 95% chance of finding at least one type 1 sample for that region class. Results are for model parameters: number potential samples per site, *nPotentialSamples*=1000; fraction of samples at a site that are type 1, given there are type 1 samples at that site, *p*=0.05; fraction of the candidate sites that have some type 1 samples, *q*=0.5; number of candidate sites within region class, Fig. 3a: *nCandidatesSites*=22; and Fig. 3b: *nCandidatesSites*=7 (Okada, 2006; Demura, 2006; Saito, 2006).



(a) **(b)**

Figure 3. Results for two different region classes: (a) rough/boulder fields, and (b) craters. Number of visited sites is plotted versus number of analyzed samples required to achieve the first success in finding a type 1 sample with a probability of at least 0.95 (blue curve), and at least 0.75 (red curve).

Expected significance based on hours devoted to surface analysis and sample collection for a proposed design reference mission. Table 2 gives the probability of not finding a type 1 sample for various values of p and the number of assumed potential samples at a site, *nPotentialSamples*. Preliminary operations concepts allocate time to analyze surface samples at different sites that range from 40 to 290 samples (see Table 1) depending on the site. Table 2 shows that depending on the site this can be too long or too short to reach a statistically significant conclusion. For example, for p=0.01, and *nPotentialSamples*=1000 these number of samples (40-290) correspond to 34% to 97% chance of observing a sample opposed to the precursor hypotheses. The results are insensitive to the number of potential samples in the Table 2. A solely timeline-based approach could be improved by incorporating a statistical modelling component.

Table 2. Estimated probability of not finding a type 1 sample for ranges of *nPotentialSamples* and *nAnalyzed* based on DRM 4b concept of operations. For p=0.25 and for all *nAnalyzed* from operation concepts gives pSite~1.0. Results show the model output is insensitive to the input parameter *nPotentialSamples*, the number potential samples per site.

p=0.01			***p=0.05***		
nPotentialSamples	***nAnalyzed***	***pSite***	***nPotentialSamples***	***nAnalyzed***	***pSite***
1000	40	0.34	1000	40	0.88
1000	50	0.40	1000	50	0.93
1000	80	0.57	1000	80	0.99
1000	100	0.65	1000	100	1.00
1000	190	0.88	1000	190	1.00
1000	290	0.97	1000	290	1.00
2000	40	0.33	2000	40	0.87
2000	50	0.40	2000	50	0.93
2000	80	0.56	2000	80	0.98
2000	100	0.64	2000	100	0.99
2000	190	0.87	2000	190	1.00
4000	40	0.33	4000	40	0.87
4000	50	0.40	4000	50	0.92
4000	80	0.56	4000	80	0.98
4000	100	0.64	4000	100	0.99

Conclusions and Discussion

The recommended number of sites to visit of a particular region class depends on the number of candidate sites, *nCandidateSites*, in that particular region class. Based on the Itokawa analog, the results suggest the number of sites visited in the rough/boulder region class (where assumed *nCandidateSites*=22) should be approximately four (Fig. 3a). The assumed number of candidate sites in the crater regions (*nCandidateSites*=7) is smaller based on the analog precursor mapping. This results in a smaller number of visited sites needed to exceed the required probability threshold (Fig. 3b).

The Hayabusa data for the number of regions of Itokawa suggests a total of 11 sites to sample. The recommended sites would be 4 sites spread over 22 rough/boulder fields, 3 sites spread over the 7 craters, 2 sites spread over the 3 regolith regions, one site at the large boulder, and one site at the unusual albedo region. In contrast if the number of sites were instead per region, not per region class, then the number of sites would be significantly larger.

The initial demonstration using Hayabusa data of Itokawa suggested the total number of sites to sample was higher than those in recent preliminary concepts of operations. While engineering and safety considerations tend to dominate the early concept of operations especially in light of the paucity of information about the potential destination, the analysis presented in this paper recommends architectures be developed that have sufficient resources to conduct statistically significant exploration operations.

The number of samples per site would be 44 samples per site to reach 90% chance of finding at least one type 1 sample at a given site (Fig. 2). Based on a 2-minute per sample analysis/examination time this would correspond to 88 minutes being devoted to sample characterization per site.

The results are insensitive to the potential samples per site, *nPotentialSamples*. This is because, for large values of *nPotentialSamples*, the model distribution converges to a binomial distribution (Rohatgi, 2003). For the same reason, the model is also somewhat insensitive to the number of candidate sites, *nCandidateSites*.

The number of samples implied by current DRM preliminary surface operations shows a range of 40-290 samples per site. The model suggests this may be too large at many sites. The model input parameters would be estimated from precursor data, and would ideally drive precursor requirements to provide measurements from which hypotheses can be formed, which in turn will drive subsequent human surface operations.

Because the model developed in this study used sampling without replacement for both samples within a site, and for visited sites within a set of candidates spread over a region class, the results are believed to be more accurate for the typically small parameter values for an asteroid mission. When coupled with the geo-statistical model of hypotheses developed herein, the accuracy can be quantified through a variety of parameters and the tradeoffs between those parameters viewed explicitly for mission planning. Sensitivity analysis over the model input parameters has been performed and will be reported separately.

Foundations. The approach described here has connections to the methodology for designing experiments first proposed by Fisher (Fisher, 1935). The approach also draws from the broad area of sampling strategies (Thompson 2002) to produce a statistically measurable sampling strategy for the exploration of asteroids with limited precursor information. The most important ideas of experimental design are applicable. In space exploration it is usually difficult to reproduce measured results exactly. Comparisons between collection strategies are difficult because of the absence of a standard control that acts as baseline. Stratified sampling strategies reduce risks such as failing to obtain a representative sample in a survey, or having a serious imbalance in a key characteristic between different sites.

Notwithstanding the above, the study had a number of limitations. Early planning for an asteroid mission has yet to select a candidate asteroid. Therefore the approach presented is limited to the Itokawa example. However, the approach is still highly relevant since precursor data would be gathered for any ultimate asteroid chosen for a human mission. Similar calculations would be performed for any candidate or set of candidates. Another limitation associated with the early planning phase is the absence of input from the science community regarding the types and quantities of samples of greatest interest for confirming or rejecting the variety of hypotheses associated with asteroids. As the mission concept further develops, it is anticipated such activities will be specified and the mission parameters will be determined. As those values become available the sampling plans for different mission architectures could easily be compared based on differences in their discovery potential.

Given existing information about asteroid regions, candidate sites, scientific hypotheses to be confirmed/denied, etc., the approach presented in this paper computes the trade space of sites visited and samples taken/site, from which mission durations can be planned to achieve scientific hypotheses verification at a required confidence level. Conversely, if one assumes a given mission duration, stops etc. one can estimate the resulting likelihood of confidently confirming/denying characterization predictions based on the number of stops, and number of samples taken. Since each visited site requires mission resources to both reach and explore a site, the model can be the basis for a tool that optimizes mission operations subject to constraints on mission resources, e.g. EVA, mobility, mission duration. Using precursor information, the system would maximize the discovery potential of the human exploration operations resulting in more efficient/cost effective exploration architectures.

Acknowledgments

The authors are indebted to John Baker of JPL for his help in defining the parameters of this study and answering our numerous questions in many interesting discussions. We thank Pam Clark for her valuable inputs and discussions about science, instruments, and operations. We thank members of HEFT II, especially James Reuther and Bill Cirillo, for background context for this study. This research was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

References

Britt, D.T., Yeomans, D., Housen, K., and Consolmagno, G., “Asteroid Densities.” EAR-A-5-DDR-ASTEROID-DENSITIES-V1.1. NASA Planetary Data System, 2002.

Demura et al, “Pole and Global Shape of 25143 Itokawa” *Science*, June 2006.

Fellows, Jack D., Rapporteur, Joseph K. Alexander, Editor. “Decadal Science Strategy Surveys, New Opportunities in Solar System Exploration”, National Research Council, 2007.

Fisher, Ronald A. *The Design of Experiments*, 1935.

HEFT Phase II. “DRM 4B 30 Day NEO ConOps”, Powerpoint Presentation, 2010.

Heiken, Grant and Jones, Eric, *On the Moon: The Apollo Journals*, Springer, 2007.

Okada, T., et. al. “X-ray Fluorescence Spectrometry of Asteroid Itokawa by Hayabusa”, *Science*, Vol. 312 no. 5778 pp. 1338-1341, June 2, 2006.
Rohatgi , Vijay K., *Statistical Inference*, Dover Publications, 2003.
Saito, J., “Detailed Images of Asteroid 25143 Itokawa from Hayabusa”, Vol. 312 no. 5778 pp. 1341-1344, Science 2 June 2006
Thomas et. al., “Eros: Shape, Topography, and Slope Processes”, *Icarus* 155, 18–37 (2002).
Thompson, Steven K. “Sampling“, Wiley 2002.
Van Schmus, W. R.; Wood, J. A. "A chemical-petrologic classification for the chondritic meteorites". *Geochimica et Cosmochimica Acta* 31: 747–765, 1967.
Weisbin, C.R., et. al, “Technical Feasibility and Relative Productivity of Alternate NASA Robotic Missions to a Lunar Dark Crater,” *IEEE Systems Journal*, Vol. 2, No. 1, pp. 120-128, March 2008.

Biography

William Lincoln received his doctoral degree from UCLA in 1995, and is a member of the High Capability Computing & Modelling Group at JPL. Dr. Lincoln has published extensively in areas of modeling and simulation, risk analysis, technology prioritization, and optimized exploration architectures.

Jeffrey H. Smith, Ph.D., is a member of the Mission and Systems Architecture Section at the JPL. He has developed decision analysis methodologies and tools for the analysis of advanced space missions and technologies for NASA. He has authored over 150 technical papers spanning a wide variety of disciplines and applications.

Charles R. Weisbin serves as Deputy Program Manager for the System Modeling and Analysis Program Office at JPL. Dr. Weisbin received the 1993 and 1999 NASA Exceptional Service Medals, and a Lifetime Achievement Award in 2004. Dr. Weisbin received the award for Best Paper published in the System Engineering Journal in 2004.