Adaptive Sampling for Multi-Robot Wide Area Prospecting

Kian Hsiang Low Geoffrey J. Gordon John M. Dolan
Pradeep Khosla

October 2005

CMU-RI-TR-05-51

October 2005

Robotics Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213

Abstract

Prospecting for *in situ* mineral resources is essential for establishing settlements on the Moon and Mars. To reduce human effort and risk, it is desirable to build robotic systems to perform this prospecting. An important issue in designing such systems is the *sampling strategy*: how do the robots choose where to prospect next? This paper argues that a strategy called *Adaptive Cluster Sampling* (ACS) has a number of desirable properties: compared to conventional strategies, (1) it reduces the total mission time and energy consumption of a team of robots, and (2) returns a higher mineral yield and more information about the prospected region by directing exploration towards areas of high mineral density, thus providing detailed maps of the boundaries of such areas. Due to the adaptive nature of the sampling scheme, it is not immediately obvious how the resulting sampled data can be used to provide an unbiased, low-variance estimate of the regional mineral density. This paper therefore investigates new mineral density estimators, which have lower error than previously-developed estimators; they are derived from the older estimators via a process called *Rao-Blackwellization*. Since the efficiency of estimators depends on the type of mineralogical population sampled, the population characteristics that favor ACS estimators are also analyzed. The ACS scheme and our new estimators are evaluated empirically in a detailed simulation of the prospecting task, and the quantitative results show that our approach can yield more minerals with less resources and provide more accurate mineral density estimates than previous methods.

Contents

1 Introduction

The establishment of large, self-sufficient lunar and Martian settlements will require an extensive use of *in situ* mineral resources. Prospecting for these resources is therefore crucial to planning these settlements [19] (e.g., site selection, processing equipment, and manufactured products). Although orbiting spacecraft can remotely survey the lunar surface for the distribution of minerals, their sensing data are limited in resolution and the types of minerals/elements sensed [13]. Hence, surface prospecting is necessary to determine the specific regions of highest abundance (in particular, geographically rare minerals and minerals not sensed by orbiters) for mining and to extract the most geologically interesting samples for detailed analysis and calibration of orbiters' data [12].

Surface prospecting can be conducted by either robots or spacesuited humans. The benefits of robot prospectors include a wider range of sensory capabilities for mineral identification, elimination of safety and life support issues, operation in harsh environments, and greater strength and endurance [6, 18]. Their deployment may increase the efficiency of sampling in large prospecting regions and relieve the humans for more sophisticated tasks such as real-time perception and planning, and detailed geologic field study.

Traditionally, conventional sampling methods such as *Raster Scanning* (RS) [1], *Simple Random Sampling* (SRS) [14], and stratified random sampling [1] have been used in prospecting with robots. The first approach acquires measurements at uniform intervals, thus incurring high sampling and travel costs to achieve adequate sampling density. The second approach selects a random sample of locations and makes measurements at each of the selected locations. However, it ignores the fact that mineral deposits are usually clustered [3, 20] and sometimes rare [19]. This results in an imprecise estimate of the mineral density in the prospecting region (i.e., large variance) [17]. Stratified random sampling requires prior knowledge of the mineral distribution for allocating the appropriate sampling effort among strata [17]. Without such information, its efficiency degrades to that of SRS. There is one other conventional sampling scheme called *Systematic Sampling* (SS) [20], which has not been utilized in robot prospecting. It will be used as a method of comparison in our paper.

This paper presents adaptive sampling techniques for wide area prospecting with a team of robots (Fig. 1). Assume that the prospecting region (Fig. 2a) is discretized into a grid of N sampling units. *Adaptive sampling* refers to sampling strategies in which the procedure for selecting units to be included in the sample depends on the mineral concentration observed during prospecting. In contrast, conventional sampling has no such dependence. The main objective of adaptive sampling is to exploit the population characteristics of mineral deposits (e.g., spatial clustering or patchiness shown in Fig. 2b) to obtain more precise estimates of the regional density than conventional strategies for a given sample size or cost.

This paper describes a specific adaptive sampling scheme known as ACS (Section 3), which has a number of desirable benefits for multi-robot wide area prospecting: (1) it returns a higher mineral yield and more information about the prospected region by directing robot exploration towards areas of high mineral density, thus providing detailed maps of the boundaries of such areas, and (2) it reduces the total mission time

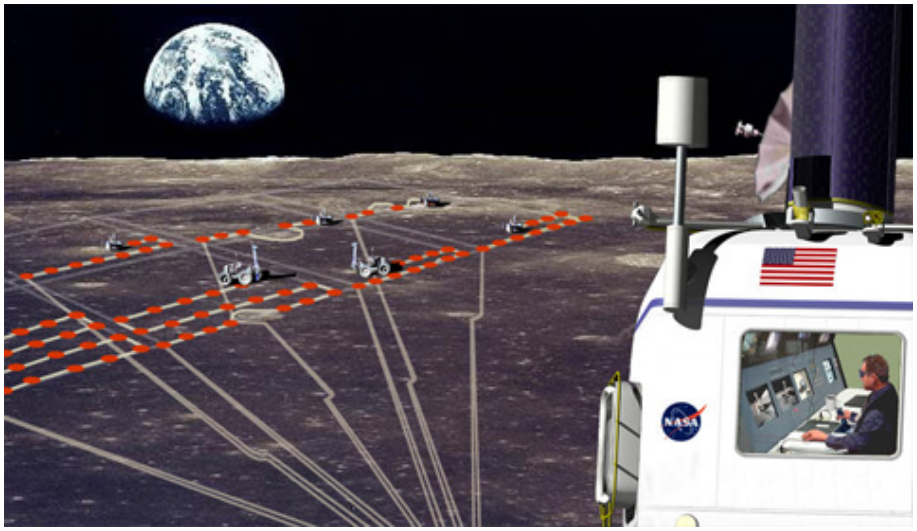


Figure 1: Multi-robot mineral prospecting task.

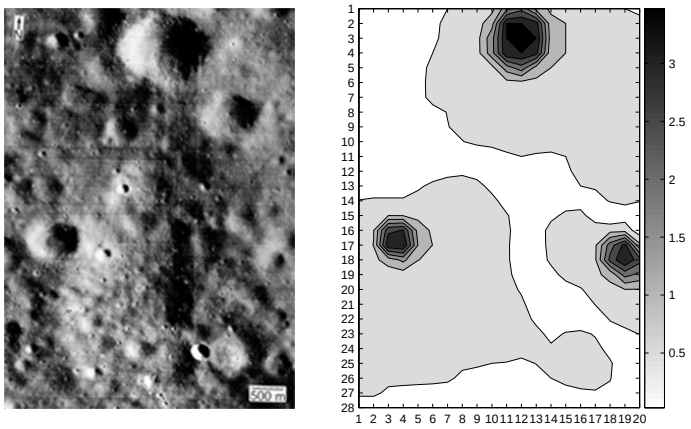
and energy consumption of the robot team (Section 7).

The adaptive nature of this scheme incurs a considerable bias in conventional estimators due to a large proportion of high mineral content data in the sample. Consequently, two unbiased estimators are proposed in [20] for the ACS strategy (Section 4). This paper investigates how the error of these estimators can be reduced through a process known as *Rao-Blackwellization*, in which the outputs of the estimators are averaged over several different ordered samples that are constructed by permuting the original sampled data. The Rao-Blackwellization procedure is elaborated in Section 5; in that section, its computational expense is also addressed, and closed-form expressions are provided for the Rao-Blackwellized estimators. Since the efficiency of estimators depends on the type of mineralogical population sampled, the population characteristics that favor ACS estimators are also analyzed (Section 6). Before discussing the ACS strategy and estimators, an overview of the multi-robot architecture will be presented first in the next section.

2 Robot Supervision Architecture

The mineral prospecting task demonstrates an application of the Robot Supervision Architecture (RSA) in our project called PROSPECT: Planetary Robots Organized for Safety and Prospecting Efficiency via Cooperative Telesupervision (http://www.ri.cmu.edu/∼prospect). This project is supported by NASA's Exploration Systems Mission Directorate. Our primary goal is to develop a general architecture for human supervision of an autonomous robot team in support of sustained, affordable, and safe space exploration.

The RSA comprises the teleoperation base and robot prospectors. The teleoperation



(a) (b)

Figure 2: (a) Lunar Orbiter photograph III-133-H2 of prospecting region north of Apollo 14 landing site (photograph courtesy of Lunar and Planetary Institute). (b) Synthetic zirconium distribution (wt$\%$) in this region [19].

base maintains a plan of the robot tours to visit the selected units to be sampled, as well as a list of sampled units and their corresponding mineral content, while each robot maintains an individual tour of its assigned units to be sampled. The base continuously receives spectrometric data from the prospecting robots, selects new sampling units based on the ACS strategy (Section 3), and replans the robot tours to visit the new and current sampling units. After all selected units have been sampled, it determines the mineral density estimates of the prospected region (Sections 4 and 5).

Our planning problem is an instance of the k-traveling salesman problem [5] where k is the number of robots. The selected sampling units can be considered as cities to be visited. We consider two different optimality criteria: minimizing the (1) total energy consumption of all robots, and the (2) maximum mission time of any robot. In general, this problem is NP-hard. So, our centralized planner at the base uses a modified minimum spanning tree heuristic proposed in [11] to obtain 2- and $2k$-competitive tour allocation for the first and second criterion respectively.

3 Adaptive Cluster Sampling

The ACS [20] scheme proceeds as follows: an initial sample of size n_1 is taken using SRS without replacement. If the observed mineral content of an initially sampled unit satisfies a certain condition C (e.g., mineral content $\geq$ predefined threshold), the unit's neighborhood is added to the sample. For every unit, its neighborhood consists of the unit and a set of "neighboring" units (e.g., top, bottom, left, and right units). If any other units in that neighborhood satisfy C, their neighborhoods are also included in the sample. This process is repeated until no more units that satisfy C are encountered.

At this stage, clusters of units are obtained. Each *cluster* contains units that satisfy C and a boundary of *edge units*. An *edge unit* is a unit that does not satisfy C but is in the neighborhood of a unit that does. The final sample of size ν consists of up to n_1 clusters. These clusters are not necessarily distinct, since two units in the initial sample that satisfy C could have been selected from the same cluster. If a unit in the initial sample does not satisfy C, it is considered to be a cluster of size one.

Let the *network* $\mathcal{A}_i$ that is generated by unit i be defined as a cluster generated by that unit with its edge units removed. A selection of any unit in $\mathcal{A}_i$ leads to the selection of all units in $\mathcal{A}_i$. Any unit that does not satisfy C is a network of size one since its selection does not lead to the inclusion of any other units. This implies that any edge unit is also a network of size one. Hence, any cluster of size larger than 1 can be decomposed into a network with units that satisfy C, and also networks (edge units) of size one that do not satisfy C. Clusters may overlap on their edge units. In contrast, networks are disjoint and form a partition of the entire population of units.

An example of an adaptive cluster sample is illustrated in Table 1. The values in this table are obtained in a simulation test run on the prospecting region in Fig. 2, which is discretized into a 28×20 grid of square sampling units (thus, the total number of units $N = 560$). The neighborhood of a unit is defined to be the top, bottom, left, and right units. The condition for sampling a unit's neighborhood is defined as $C = (y \geq 1.0 \text{ wt}\%)$ where y is the observed mineral content of a sampling unit. With an initial sample size n_1 of 80, the ACS scheme results in a final sample size ν of 150. The boxed values correspond to units from the initial sample. The lightly and darkly shaded units correspond to network and edge units respectively. 3 networks of size larger than 1 can be observed in the sample (Table 1). Notice that the leftmost network is intersected 4 times by the initial sample while the other two networks are each intersected once.

A noteworthy aspect of ACS is that given a fixed ν, the travel cost of adding a cluster or network of "neighboring" units in ACS is usually lower than that of adding units selected at random using SRS in the prospecting region. This is demonstrated empirically in Section 7.

Since the ACS scheme results in a large proportion of high mineral content data in the sample, it will incur a considerable bias with the conventional sample mean estimator $\bar{\mu} = \nu^{-1} \sum_{i=1}^{\nu} y_i$. Kriging (or Gaussian process regression) [4] is a more sophisticated alternative but will be similarly biased. For example, the true population mean μ is 0.648 for the zirconium distribution in Fig. 2b. However, for the ACS example in Table 1, $\bar{\mu} = 1.070$ with $\widehat{\text{var}}[\bar{\mu}] = 0.004139$ (i.e., standard error of 0.064), which clearly overestimates μ. Hence, unbiased estimators are needed for the ACS scheme. Two of these are presented in the next section.

4 Unbiased ACS Estimators

4.1 Modified Horvitz-Thompson Estimator

The first ACS estimator is modified from the Horvitz-Thompson (HT) estimator [8]. Let $\mathcal{B}_i$ be the set of units in the ith network and m_i be the number of units in $\mathcal{B}_i$. Note

Table 1: ACS example. Please refer to Section 3 for its description.

									0.85	1.83	2.35	1.31	0.90	0.85					
		0.00						0.80	1.31	3.50	3.50	3.21	1.16	1.00	0.90			0.70	
				0.10				0.82	1.83	3.50	3.50	3.50	1.25	1.00	0.96				
			0.00					0.81	1.80	3.45	3.50	3.18	1.13	1.00	0.96			0.84	0.80
	0.00								0.85	1.75	2.30	1.30	1.00	0.96			0.86		
0.00	0.00	0.00						0.76	0.80	0.85	0.87	0.87	0.86					0.64	
					0.59					0.74									
	0.00					0.49	0.58	0.63						0.68					
		0.00							0.29			0.48		0.50					
										0.02									0.76
									0.22			0.06				0.56			
		0.60						0.81	0.76							0.31			
	0.80	1.00	0.98	0.86												0.20			
	0.98	2.88	3.03	1.31	0.90												0.83	0.81	0.20
0.79	1.01	3.40	3.50	1.50	0.99		0.80								0.86	0.94	2.25	3.19	1.12
	0.88	1.21	1.63	1.01	0.90				0.64	0.60						0.95	2.67	3.50	2.70
		0.73	0.74	0.74		0.75					0.20	0.00				0.93	1.10	2.20	1.84
			0.74					0.84									0.90	0.99	0.98
											0.67		0.00		0.00	0.53			
			0.77	0.85					0.87							0.00			
												0.68			0.62		0.10	0.25	
			0.79					0.86				0.70						0.00	
			0.79				0.70					0.52			0.80				
			0.70											0.57					
														0.42				0.45	
																		0.44	

that $\mathcal{B}_i$ is defined in the same way as network $\mathcal{A}_i$ in Section 3 except that its index i refers to the network label rather than the unit label. The probability that the initial sample intersects network $\mathcal{B}_i$ is

$$\pi_i \stackrel{\text{def}}{=} 1 - \binom{N - m_i}{n_1} \Big/ \binom{N}{n_1} . \tag{1}$$

The total mineral content of the prospecting region can be written as the sum of the mineral contents of the individual networks. So, the average mineral content is

$$\mu = \frac{1}{N} \sum_{i=1}^{K} y_i^*$$

where y_i^* is the total mineral content of the ith network and K is the total number of distinct networks in the population.

μ cannot be computed directly due to the unknown y_i^*'s for unsampled networks. So, to form an unbiased estimator of μ, each term in the sum can be multiplied by I_i/π_i, where I_i is an indicator variable of value 1 if the initial sample intersects $\mathcal{B}_i$, and

0 otherwise. The expected value of I_i/π_i is 1 for sampled networks, so our estimator is unbiased; since I_i is 0 for unsampled networks, information about these networks are not needed to calculate our estimator. Applying this trick yields the *modified HT estimator* of μ:

$$\hat{\mu}_{HT} = \sum_{i=1}^{K} \frac{y_i^* I_i}{N\pi_i} = \sum_{i=1}^{\kappa} \frac{y_i^*}{N\pi_i} \tag{2}$$

where κ is the number of distinct networks intersected by the initial sample.

For practical use of the HT estimator, it is important to be able to estimate its variance from the sample. There is a simple closed-form formula which can be used for this purpose. π_i has been defined to be the probability that the initial sample intersects the ith network. Define π_{jk} to be the probability that the initial sample intersects both the jth and kth networks. If $j = k$, then $\pi_{jk} = \pi_j$. Otherwise, to compute π_{jk}, notice that the probability that the initial sample intersects neither network j nor network k is

$$\mathrm{P}(I_j \neq 1 \cap I_k \neq 1) = \binom{N - m_j - m_k}{n_1} \Big/ \binom{N}{n_1} .$$

So, the probability that the initial sample intersects either jth or kth network is $1-\mathrm{P}(I_j \neq 1 \cap I_k \neq 1)$, and

$$\pi_{jk} = \pi_j + \pi_k - (1 - \mathrm{P}(I_j \neq 1 \cap I_k \neq 1)) .$$

Since $\hat{\mu}_{HT}$ is a sum of several terms, its variance can be derived by taking the sum of covariances between these terms:

$$\begin{aligned} \mathrm{var}[\hat{\mu}_{HT}] = & \sum_{j=1}^{K}\sum_{k=1}^{K} \mathrm{cov}[\frac{y_j^* I_j}{N\pi_j}, \frac{y_k^* I_k}{N\pi_k}] \\ = & \sum_{j=1}^{K}\sum_{k=1}^{K} \frac{y_j^*}{N\pi_j}\frac{y_k^*}{N\pi_k}\mathrm{cov}[I_j, I_k] . \end{aligned} \tag{3}$$

(3) cannot be computed from the sample data since not all the networks in the population are necessarily sampled. So, to obtain an unbiased estimator of the variance, we can use a similar trick as before: each term is multiplied by $I_j I_k/\pi_{jk}$ (which has an expected value of 1 for sampled networks) to get

$$\begin{aligned} \widehat{\mathrm{var}}[\hat{\mu}_{HT}] = & \sum_{j=1}^{K}\sum_{k=1}^{K} \frac{y_j^* I_j}{N\pi_j}\frac{y_k^* I_k}{N\pi_k}\frac{\mathrm{cov}[I_j, I_k]}{\pi_{jk}} \\ = & \frac{1}{N^2}\left[\sum_{j=1}^{\kappa}\sum_{k=1}^{\kappa} \frac{y_j^* y_k^*}{\pi_{jk}}\left(\frac{\pi_{jk}}{\pi_j\pi_k} - 1\right)\right] . \end{aligned} \tag{4}$$

The second equality follows because $\mathrm{cov}[I_j, I_k]$ is $\pi_{jk} - \pi_j\pi_k$.

For the ACS example in Table 1, $\hat{\mu}_{HT} = 0.665$ and $\widehat{\mathrm{var}}[\hat{\mu}_{HT}] = 0.001835$ (i.e., standard error of 0.043) by using (2) and (4) respectively.

4.2 Modified Hansen-Hurwitz Estimator

The second ACS estimator is modified from the Hansen-Hurwitz (HH) estimator [7]. In the previous section, we mention that the total mineral content of the prospecting region is the sum of the mineral contents of the individual networks. The mineral content of each network can be written as the average mineral content of all units in this network summed over its number of network units. So, the average mineral content of the prospecting region can also be expressed as

$$\mu = \frac{1}{N}\sum_{i=1}^{N} w_i$$

where w_i is the average mineral content of the network $\mathcal{A}_i$ containing unit i.

μ cannot be computed directly due to the unknown w_i's for unsampled networks. Using the same trick as in the previous section, an unbiased estimator of μ can be formed by multiplying each term in the sum with NJ_i/n_1, where J_i is an indicator variable of value 1 if unit i is included in the initial sample, and 0 otherwise. The expected value of NJ_i/n_1 is 1 for initial sample units, so our estimator is unbiased; since J_i is 0 for units not in the initial sample, information about these units is not needed to calculate our estimator. Applying this trick yields the *modified HH estimator* of μ:

$$\hat{\mu}_{HH} = \frac{1}{n_1}\sum_{i=1}^{N} w_i J_i = \frac{1}{n_1}\sum_{i=1}^{n_1} w_i \tag{5}$$

Note that $\hat{\mu}_{HH}$ can be interpreted as the conventional sample mean obtained using SRS of size n_1 from a population of w_i values rather than y_i values. So, using the theory of SRS [20],

$$\text{var}[\hat{\mu}_{HH}] = \frac{N-n_1}{Nn_1(N-1)}\sum_{i=1}^{N}(w_i-\mu)^2 \tag{6}$$

with unbiased estimator

$$\widehat{\text{var}}[\hat{\mu}_{HH}] = \frac{N-n_1}{Nn_1(n_1-1)}\sum_{i=1}^{n_1}(w_i-\hat{\mu}_{HH})^2\,. \tag{7}$$

For the ACS example in Table 1, $\hat{\mu}_{HH} = 0.624$ and $\widehat{\text{var}}[\hat{\mu}_{HH}] = 0.002802$ (i.e., standard error of 0.053) by using (5) and (7) respectively.

5 Improved Unbiased Rao-Blackwellized ACS Estimators

An estimator $t(D_o)$ of a population characteristic μ is a function t which maps our observed data D_o to an estimate of μ. Saying that μ is a population characteristic means there is a parameter vector $\boldsymbol{\theta}$ which completely describes the distribution of our population, and $\mu = \mu(\boldsymbol{\theta})$ is a function of this parameter vector.

In our setting, D_o is an ordered list of pairs $\langle i_s, y_{i_s} \rangle$ where i_s is the unit sampled at step s and y_{i_s} is its mineral content. The population characteristic of interest μ is the average mineral content of the sampling region. The population parameter is $\boldsymbol{\theta} = \langle y_1, \ldots, y_N \rangle$, which is the vector of true mineral contents for all units in the population. The estimators of μ that we are interested in are $\hat{\mu}_{HT}$ and $\hat{\mu}_{HH}$.

To evaluate an estimator $t(D_o)$, its distribution conditioned on a possible value of θ can be examined. Good estimators have low mean-squared errors, i.e., the distribution P($t(D_0) - \mu|\boldsymbol{\theta}$) is concentrated around 0. In this section, we will describe a way to reduce the mean-squared errors of the estimators $\hat{\mu}_{HT}$ and $\hat{\mu}_{HH}$.

Rao-Blackwellization is a procedure that allows us to reduce the mean-squared error of an arbitrary estimator $t(D_o)$. The improved estimator is E($t(D_o)|\mathcal{D}$) where $\mathcal{D}$ is a reduced description of our data that omits some redundant information. In particular, $\mathcal{D}$ is defined as a *statistic* if it is a function of our data D_o, and $\mathcal{D}$ is defined as a *sufficient statistic* if it contains all relevant information in D_o about $\boldsymbol{\theta}$, i.e., P($D_o|\mathcal{D}, \boldsymbol{\theta}$) = P($D_o|\mathcal{D}$).

Given these definitions, Rao-Blackwellization is the process of computing E($t(D_o)|\mathcal{D}$) when $\mathcal{D}$ is a sufficient statistic. In our case, $\mathcal{D}$ is set to be the *unordered* set of distinct, labeled observations, i.e., $\mathcal{D} = \{\langle i, y_i \rangle | \; i \in \mathcal{S}\}$ where $\mathcal{S}$ is the set of distinct unit labels in our data sample [20].

The following theorem, adapted from the Rao-Blackwell theorem [2], justifies the use of the Rao-Blackwellized (RB) estimator:

Theorem 1 *Let $t = t(D_o)$ be a (not necessarily unbiased) estimator of μ. Define $t_{\mathcal{D}} = E[t|\mathcal{D}]$. Then*
(a) $t_{\mathcal{D}}$ is an estimator;
(b) $E[t_{\mathcal{D}}] = E[t]$;
(c) MSE$[t_{\mathcal{D}}] \leq$ MSE$[t]$ with strict inequality for all $\boldsymbol{\theta}$ such that $P_{\boldsymbol{\theta}}(t \neq t_{\mathcal{D}}) > 0$.

Corollary 1 *If t is unbiased,*

$$\begin{aligned} var[t_{\mathcal{D}}] = &\quad var[t] - E_{\mathcal{D}}E[(t - t_{\mathcal{D}})^2|\mathcal{D}] \\ = &\quad var[t] - E_{\mathcal{D}}\{var[t|\mathcal{D}]\} \; . \end{aligned} \tag{8}$$

The proofs of Theorem 1 and Corollary 1 are provided in Appendices A and B respectively. From (8), var$[t_{\mathcal{D}}] \leq$ var$[t]$ since the variance reduction term E$_{\mathcal{D}}\{$var$[t|\mathcal{D}]\} \geq 0$.

Rao-Blackwellization does nothing if $g(D_o)$ is already a function of $\mathcal{D}$. On the other hand, it achieves the largest possible reduction in variance when $\mathcal{D}$ is a *minimal sufficient statistic*. A minimal sufficient statistic is one that reduces D_o as much as possible without losing information about $\boldsymbol{\theta}$:

Definition 1 *A sufficient statistic $\mathcal{D} = g(D_o)$ is* minimal sufficient *for $\boldsymbol{\theta}$ if, for any other sufficient statistic $\mathcal{D}' = g'(D_o)$, $\mathcal{D}$ is a function of $\mathcal{D}'$.*

In our case, $\mathcal{D}$ is minimal sufficient, and $\hat{\mu}_{HT}$ and $\hat{\mu}_{HH}$ are not functions of $\mathcal{D}$; they depend on the order of selection. To see this, consider a small population of four units with $\boldsymbol{\theta} = [0.1, 0.5, 1, 2]^T$. The condition is $y \geq 1$. The initial sample size n_1 is 2. The initial samples ($\langle 2, 0.5 \rangle, \langle 3, 1 \rangle$) and ($\langle 3, 1 \rangle, \langle 4, 2 \rangle$) give the same final unordered

sample $\mathcal{D} = \{\langle 2, 0.5\rangle, \langle 3, 1\rangle, \langle 4, 2\rangle\}$. Note that the edge unit $\langle 2, 0.5\rangle$ is included in the first initial sample but not the second one. This implies the edge unit will be involved in computing $\hat{\mu}_{HT}$ or $\hat{\mu}_{HH}$ only in the first sample. Hence, either estimator will produce different estimates for the two samples.

In order to Rao-Blackwellize $\hat{\mu}_{HT}$ and $\hat{\mu}_{HH}$, we will need several notations. Let $G = \binom{\nu}{n_1}$ be the number of combinations of n_1 distinct initial sample units from the ν units in the final sample and let these combinations be indexed by the label g where $g = 1, 2, \ldots, G$. Let τ_g be the value of an estimator t when the initial sample consists of combination g, I_g be an indicator variable of value 1 if the gth combination can result in $\mathcal{D}$ (i.e., is *compatible* with $\mathcal{D}$), and 0 otherwise. The number of compatible combinations is then $\xi = \sum_{g=1}^{G} I_g$. It follows that P($t = \tau_g|\mathcal{D}$) = $1/\xi$ for all compatible g. So, the improved RB estimator is

$$t_{RB} = \mathrm{E}[t|\mathcal{D}] = \frac{1}{\xi}\sum_{g=1}^{G}\tau_g I_g = \frac{1}{\xi}\sum_{g=1}^{\xi}\tau_g \ . \tag{9}$$

The variance of t_{RB} is obtained using (8) where $t_{\mathcal{D}} = t_{RB}$. The unbiased estimator of var[t_{RB}] is then

$$\widehat{\mathrm{var}}[t_{RB}] = \widehat{\mathrm{var}}[t] - \mathrm{var}[t|\mathcal{D}] = \widehat{\mathrm{var}}[t] - \frac{1}{\xi}\sum_{g=1}^{\xi}(\tau_g - t_{RB})^2 \ . \tag{10}$$

Since (9) and (10) are based on samples compatible with $\mathcal{D}$, naively, the ξ compatible samples have to be identified from the G combinations and their corresponding ξ estimators have to be evaluated. ξ and G can be potentially large, which would render the RB method computationally infeasible. However, in the next two subsections, closed-form expressions for the RBHT and RBHH estimators [16] will be described.

5.1 Rao-Blackwellized HT Estimator

The reason that the HT estimator yields different values with different compatible samples is each compatible sample intersects a different combination of the edge units in the final sample $\mathcal{D}$ (in contrast, all networks other than the edge units must be included in every compatible sample). So, some of the indicator variables in (2) will have different values in different compatible samples.

The closed-form expression for the RBHT estimator is based on the observation that we can analytically compute the expectation of each of the indicator variables in (2) given a randomly selected compatible sample. This expectation will be 1 for all networks of size greater than 1 and for all networks of size 1 that are not edge units. For networks that are edge units, the expectation will be strictly positive and less than 1.

As we will see below, computing these expectations requires evaluating several binomial coefficients. The number of binomial coefficients is exponential in the number of networks of size greater than 1. So, computing the RBHT estimator will be efficient if relatively few networks of size larger than 1 are intersected by the initial sample.

This assumption is reasonable because a prospecting region typically contains only a few major mineral deposits.

Since the modified HT estimator is formulated based on networks, its RB version will be derived likewise. Let $\mathcal{E}$ be the set of network labels in the sample. From (2), $\hat{\mu}_{HT}$ for the gth combination can be written as

$$\hat{\mu}_{HT}^{g} = \sum_{i \in \mathcal{E}} \frac{y_i^*}{N\pi_i} I_{gi} \tag{11}$$

where I_{gi} is an indicator variable of value 1 if the gth combination contains at least one unit from $\mathcal{B}_i$, and 0 otherwise.

By substituting (11) into (9), we get the RBHT estimator of μ:

$$\hat{\mu}_{RBHT} = \frac{1}{\xi} \sum_{g=1}^{\xi} \sum_{i \in \mathcal{E}} \frac{y_i^*}{N\pi_i} I_{gi} = \sum_{i \in \mathcal{E}} \frac{y_i^*}{N\pi_i} \sum_{g=1}^{\xi} \frac{I_{gi}}{\xi} \tag{12}$$

where $\sum_{g=1}^{\xi} I_{gi}$ is the number of combinations containing at least one unit from $\mathcal{B}_i$. Clearly, ξ and $\sum_{g=1}^{\xi} I_{gi}$ in (12) have to be evaluated in order to obtain the closed-form expression of $\hat{\mu}_{RBHT}$. This can be achieved by evaluating them based on the different types of network in the sample.

The sample $\mathcal{D}$ can be partitioned into three different types of network: (a) networks of size larger than 1, (b) networks that are edge units, and (c) networks of size one that are not edge units. More formally, $\mathcal{E} = \mathcal{F}_1 \cup \mathcal{F}_2 \cup \mathcal{F}_3$ where (a) $\mathcal{F}_1 = \{i \in \mathcal{E} |\ m_i > 1\}$, (b) $\mathcal{F}_2 = \{i \in \mathcal{E} |\ \mathcal{B}_i \text{ is an edge unit}\}$, and (c) $\mathcal{F}_3 = \{i \in \mathcal{E} - \mathcal{F}_1 - \mathcal{F}_2\}$.

ξ can be determined as follows: every network in $\mathcal{F}_3$ must be intersected by the initial sample and is thus allocated one initial sample unit each. The remaining $n_1' = n_1 - |\mathcal{F}_3|$ initial sample units can be chosen from $\nu' = \nu - |\mathcal{F}_3|$ final sample units in $\binom{\nu'}{n_1'}$ ways. From these $\binom{\nu'}{n_1'}$ ways, combinations that are not compatible with $\mathcal{D}$ have to be removed to get ξ. These incompatible combinations are the ones that contain no units from at least one of the networks in $\mathcal{F}_1$. Formally, if $\mathcal{C}_i$ is the set of combinations containing no units from $\mathcal{B}_i$, $\cup_{i \in \mathcal{F}_1} \mathcal{C}_i$ is the set of incompatible combinations that contain no units from at least one of the networks in $\mathcal{F}_1$. Using the inclusion-exclusion identity for the cardinality of set union,

$$\begin{aligned}
\left| \bigcup_{i \in \mathcal{F}_1} \mathcal{C}_i \right| &= \sum_{i \in \mathcal{F}_1} |\mathcal{C}_i| - \sum_{i,j \in \mathcal{F}_1} |\mathcal{C}_i \cap \mathcal{C}_j| + \ldots + (-1)^{|\mathcal{F}_1 - 1|} \left| \bigcap_{i \in \mathcal{F}_1} \mathcal{C}_i \right| \\
&= \sum_{i \in \mathcal{F}_1} \binom{\nu' - m_i}{n_1'} - \sum_{i,j \in \mathcal{F}_1} \binom{\nu' - m_i - m_j}{n_1'} \\
&\quad + \ldots + (-1)^{|\mathcal{F}_1 - 1|} \binom{\nu' - \sum_{i \in \mathcal{F}_1} m_i}{n_1'}.
\end{aligned}$$

Then the number of compatible combinations is simply

$$\xi = \binom{\nu'}{n_1'} - \left| \bigcup_{i \in \mathcal{F}_1} \mathcal{C}_i \right|. \tag{13}$$

Since all compatible combinations contain at least one unit from each network in $\mathcal{F}_1 \cup \mathcal{F}_3$, $\sum_{g=1}^{\xi} I_{gi} = \xi$ for $i \in \mathcal{F}_1 \cup \mathcal{F}_3$. This takes care of the cases (a) and (c). However, each edge unit in $\mathcal{F}_2$ is not necessarily included in every combination. To include an edge unit in a combination, one initial sample unit has to be allocated to the edge unit, which results in $\nu' - 1$ units in $\mathcal{D}$ and $n_1' - 1$ initial sample units. So, if we let ξ_1 be the number of combinations containing the edge unit $\mathcal{B}_i$ (i.e., $\sum_{g=1}^{\xi} I_{gi} = \xi_1$ for $i \in \mathcal{F}_2$), ξ_1 is similar to ξ except that both terms in every binomial coefficient are reduced by one. Substituting ξ and ξ_1 into (12),

$$\hat{\mu}_{RBHT} = \sum_{i \in \mathcal{F}_1 \cup \mathcal{F}_3} \frac{y_i^*}{N\pi_i} + \frac{\xi_1}{\xi} \sum_{i \in \mathcal{F}_2} \frac{y_i^*}{N\pi_i} \,. \tag{14}$$

From (14), $\hat{\mu}_{RBHT}$ is similar to $\hat{\mu}_{HT}$ except that all edge units in $\mathcal{F}_2$ are now involved but each edge unit term is weighted by ξ_1/ξ. Hence, $\hat{\mu}_{RBHT}$ is improved from $\hat{\mu}_{HT}$ by using all information from the edge units. This is also the rationale for improving the variance estimate of $\hat{\mu}_{RBHT}$, which will be described next.

It is shown in Appendix C that

$$\text{var}[\hat{\mu}_{HT}|\mathcal{D}] = \frac{1}{(n_1\xi)^2} \left((\xi_1\xi - \xi_1^2) \sum_{i \in \mathcal{F}_2} y_i^{*2} + 2(\xi_2\xi - \xi_1^2) \sum_{i \in \mathcal{F}_2} \sum_{j<i} y_i^* y_j^* \right) \tag{15}$$

where ξ_2 is the number of combinations containing any two edge units in $\mathcal{F}_2$. Two initial sample units are allocated to two edge units, resulting in $\nu' - 2$ units in $\mathcal{D}$ and $n_1' - 2$ initial sample units. So, ξ_2 is similar to ξ except that both terms in every binomial coefficient are reduced by two. From (10), the unbiased estimator of var$[\hat{\mu}_{RBHT}]$ is

$$\widehat{\text{var}}[\hat{\mu}_{RBHT}] = \widehat{\text{var}}[\hat{\mu}_{HT}] - \text{var}[\hat{\mu}_{HT}|\mathcal{D}] \tag{16}$$

where $\widehat{\text{var}}[\hat{\mu}_{HT}]$ and var$[\hat{\mu}_{HT}|\mathcal{D}]$ are previously determined using (4) and (15) respectively. It can be observed from (15) that the variance reduction term var$[\hat{\mu}_{HT}|\mathcal{D}]$ depends solely on the y-values of the edge units. Hence, the improvement in the RBHT estimator becomes smaller as the y-values of the edge units tend to zero. For the ACS example in Table 1, $\hat{\mu}_{RBHT} = 0.656$ and $\widehat{\text{var}}[\hat{\mu}_{RBHT}] = 0.001574$ (i.e., standard error of 0.040) by using (14) and (16) respectively. The standard error is only reduced by 0.003, which is expected since the y-values of the edge units are less than 1.0 (Section 3).

Note that ξ, ξ_1, and ξ_2 each requires $2^{|\mathcal{F}_1|}$ binomial coefficients to be evaluated. Hence, the expressions for $\hat{\mu}_{RBHT}$ and $\widehat{\text{var}}[\hat{\mu}_{RBHT}]$ are computationally efficient if $|\mathcal{F}_1|$ is small, i.e., relatively few networks of size larger than 1 are intersected by the initial sample.

5.2 Rao-Blackwellized HH Estimator

The derivation of the closed-form expression for the RBHH estimator follows the same notion as before: we evaluate expectations of indicator variables by counting how often

individual units are present in a compatible sample. In the RBHT estimator, the indicator variables for networks in $\mathcal{F}_1 \cup \mathcal{F}_3$ are always 1, and we only need to calculate the expectations for the networks in $\mathcal{F}_2$. The RBHH estimator, on the other hand, uses indicator variables of individual units. As before, all units in $\mathcal{F}_3$ will be present in every compatible sample and so, their indicator variables will always be 1. But, the indicator variables for the units in networks in $\mathcal{F}_1 \cup \mathcal{F}_2$ will be strictly between 0 and 1: both edge units and units in networks of size greater than 1 are not necessarily included in all of the compatible samples.

Since the modified HH estimator is formulated based on the w-value for each unit, its RB version will be derived likewise. The set $\mathcal{S}$ of unit labels in the sample $\mathcal{D}$ can be partitioned into three different types of units: (a) units in networks of size larger than 1, (b) edge units, and (c) non-edge units in networks of size one. So, $\mathcal{S} = \mathcal{H}_1 \cup \mathcal{H}_2 \cup \mathcal{H}_3$ where (a) $\mathcal{H}_1 = \{i \in \mathcal{S} |$ unit $i \in \mathcal{B}_j, j \in \mathcal{F}_1\}$, (b) $\mathcal{H}_2 = \{i \in \mathcal{S} |$ unit $i \in \mathcal{B}_j, j \in \mathcal{F}_2\}$, and (c) $\mathcal{H}_3 = \{i \in \mathcal{S} |$ unit $i \in \mathcal{B}_j, j \in \mathcal{F}_3\}$. Note that $|\mathcal{H}_2| = |\mathcal{F}_2|$ and $|\mathcal{H}_3| = |\mathcal{F}_3|$.

Let J_{gi} be an indicator variable of value 1 if the gth combination contains unit i, and 0 otherwise. From (5), $\hat{\mu}_{HH}$ for the gth combination can be expressed as

$$\hat{\mu}^g_{HH} = \frac{1}{n_1} \sum_{i \in \mathcal{S}} w_i J_{gi} = \frac{1}{n_1} \left(\sum_{i \in \mathcal{H}_1 \cup \mathcal{H}_2} w_i J_{gi} + \sum_{i \in \mathcal{H}_3} w_i \right) \tag{17}$$

since all unit labels in $\mathcal{H}_3$ must be in every gth combination. By substituting (17) into (9), the RBHH estimator of μ is

$$\begin{aligned} \hat{\mu}_{RBHH} &= \frac{1}{\xi} \sum_{g=1}^{\xi} \frac{1}{n_1} \left(\sum_{i \in \mathcal{H}_1 \cup \mathcal{H}_2} w_i J_{gi} + \sum_{i \in \mathcal{H}_3} w_i \right) \\ &= \frac{1}{n_1} \left(\frac{1}{\xi} \sum_{i \in \mathcal{H}_1 \cup \mathcal{H}_2} w_i \sum_{g=1}^{\xi} J_{gi} + \sum_{i \in \mathcal{H}_3} w_i \right) \end{aligned} \tag{18}$$

where $\sum_{g=1}^{\xi} J_{gi} = \xi_i$ is the number of combinations containing unit i. ξ can be determined using (13) in the previous section. To obtain ξ_i, one initial sample unit is allocated to unit i and each unit in $\mathcal{H}_3$, resulting in $\nu' - 1 = \nu - |\mathcal{H}_3| - 1$ units in $\mathcal{D}$ and $n_1' - 1 = n_1 - |\mathcal{H}_3| - 1$ initial sample units. ξ_i is then $\binom{\nu'-1}{n_1'-1}$ ways reduced by the number of combinations that contain no unit label from at least one of the networks in $\mathcal{F}_1$. For $i \in \mathcal{H}_1$, ξ_i is similar to ξ_1 except that we ignore the binomial coefficients involving $j \in \mathcal{F}_1$ where unit $i \in \mathcal{B}_j$. For $i \in \mathcal{H}_2$, $\xi_i = \xi_1$. Substituting ξ_i into (18),

$$\hat{\mu}_{RBHH} = \frac{1}{n_1} \left(\frac{1}{\xi} \sum_{i \in \mathcal{H}_1 \cup \mathcal{H}_2} w_i \xi_i + \sum_{i \in \mathcal{H}_3} w_i \right) . \tag{19}$$

From (19), $\hat{\mu}_{RBHH}$ differs from $\hat{\mu}_{HH}$ by involving all network units in $\mathcal{H}_1$ and edge units, but each term with these units is weighted by ξ_i/ξ. Hence, $\hat{\mu}_{RBHH}$ is improved from $\hat{\mu}_{HH}$ by using all information from the network units in $\mathcal{H}_1$ and edge units, which is also the case for improving the variance estimate of $\hat{\mu}_{RBHH}$. Similar to deriving

var$[\hat{\mu}_{HT}|\mathcal{D}]$ in Appendix C,

$$\begin{aligned}\text{var}[\hat{\mu}_{HH}|\mathcal{D}] = \quad & \frac{1}{(n_1\xi)^2}\Big(\sum_{i\in\mathcal{H}_1\cup\mathcal{H}_2} (\xi_i\xi - \xi_i^2)w_i^2 \\ & +2\sum_{i\in\mathcal{H}_1\cup\mathcal{H}_2}\sum_{j<i}(\xi_{ij}\xi - \xi_i\xi_j)w_iw_j\Big)\end{aligned} \tag{20}$$

where ξ_{ij} is the number of combinations containing units i and j in $\mathcal{H}_1 \cup \mathcal{H}_2$, that is,

$$\xi_{ij} = \begin{cases} \binom{\nu'-2}{n_1'-2} - \sum_{x\neq k,l}\binom{\nu'-m_x-2}{n_1'-2} & \text{if unit } i\in\mathcal{B}_k \wedge \text{unit } j\in\mathcal{B}_l, \\ +\sum_{x,y\neq k,l}\binom{\nu'-m_x-m_y-2}{n_1'-2} & k,l\in\mathcal{F}_1, \\ +\ldots+(-1)^{|\mathcal{F}_1|}\binom{\nu'-\sum m_x-2}{n_1'-2} & \\ \binom{\nu'-2}{n_1'-2} - \sum_{x\neq k}\binom{\nu'-m_x-2}{n_1'-2} & \text{if (unit } i\in\mathcal{B}_k, k\in\mathcal{F}_1 \\ +\sum_{x,y\neq k}\binom{\nu'-m_x-m_y-2}{n_1'-2} & \wedge\, j\in\mathcal{H}_2)\vee(\text{units } i,j\in\mathcal{B}_k, \\ +\ldots+(-1)^{|\mathcal{F}_1|}\binom{\nu'-\sum m_x-2}{n_1'-2} & k\in\mathcal{F}_1), \\ \xi_2 & \text{if } i,j\in\mathcal{H}_2 \end{cases} \tag{21}$$

where $x, y \in \mathcal{F}_1$. From (10), the unbiased estimator of var$[\hat{\mu}_{RBHH}]$ is

$$\widehat{\text{var}}[\hat{\mu}_{RBHH}] = \widehat{\text{var}}[\hat{\mu}_{HH}] - \text{var}[\hat{\mu}_{HH}|\mathcal{D}] \tag{22}$$

where $\widehat{\text{var}}[\hat{\mu}_{HH}]$ and var$[\hat{\mu}_{HH}|\mathcal{D}]$ are previously determined using (7) and (20) respectively. By comparing (15) and (20), we can observe that var$[\hat{\mu}_{HH}|\mathcal{D}] >$ var$[\hat{\mu}_{HT}|\mathcal{D}]$. If the y-values of the edge units tend to zero, var$[\hat{\mu}_{HH}|\mathcal{D}]$ depends more on the w-values of the network units in $\mathcal{H}_1$. Thus, the improvement of the RBHH estimator is always greater than that of the RBHT estimator. For the ACS example in Table 1, $\hat{\mu}_{RBHH} = 0.648$ and $\widehat{\text{var}}[\hat{\mu}_{RBHT}] = 0.000720$ (i.e., standard error of 0.027) by using (19) and (22) respectively. The standard error is reduced by 0.026, which is more than the reduction of 0.003 for the RBHT estimator (Section 5.1). Similar to the complexity analysis of RBHT estimator, the expressions for $\hat{\mu}_{RBHH}$ and $\widehat{\text{var}}[\hat{\mu}_{RBHH}]$ are computationally efficient if $|\mathcal{F}_1|$ is small.

6 Efficiency Analysis of ACS Estimators

The efficiency of ACS over SRS depends on the type of mineralogical population being sampled. In particular, $\hat{\mu}_{HH}$ is more efficient than the conventional sample mean $\hat{\mu}$ for SRS if var$[\hat{\mu}_{HH}] <$ var$[\hat{\mu}]$. Using the theory of SRS [20],

$$\text{var}[\hat{\mu}] = \frac{N-\nu}{N\nu(N-1)}\sum_{i=1}^{N}(y_i-\mu)^2 \tag{23}$$

The total sum of squared difference between y_i and μ in (23) can be partitioned into within-network and between-network components:

$$\sum_{i=1}^{N}(y_i-\mu)^2 = \sum_{i=1}^{N}(y_i-w_i)^2 + \sum_{i=1}^{N}(w_i-\mu)^2 \tag{24}$$

Using (6), (23), and (24), var$[\hat{\mu}_{HH}] <$ var$[\hat{\mu}]$ if and only if

$$\Big(1 - \frac{n_1}{\nu}\Big) \sum_{i=1}^{N} (y_i - \mu)^2 < \Big(1 - \frac{n_1}{N}\Big) \sum_{i=1}^{N} (y_i - w_i)^2 \tag{25}$$

It can be observed from (25) that $\hat{\mu}_{HH}$ is more efficient than $\hat{\mu}$ if (1) the within-network variance of the population (rightmost term) is sufficiently high, (2) the final sample size ν is not much larger than the initial sample size n_1 for $\hat{\mu}_{HH}$ so that $1 - n_1/\nu$ is small, and (3) $n_1 \ll N$ so that $1 - n_1/N$ is large. However, conditions 2 and 3 can oppose condition 1 because a small difference between initial and final sample size, and a small initial sample size usually mean small within-network variance. So, ACS with $\hat{\mu}_{HH}$ works best with networks that are small enough to restrict the final sample size but large enough for the within-network variance to represent the population variance reasonably. This means even though drastically lowering the threshold for condition C can increase the within-network variance and improve condition 1, it also increases the final sample size tremendously and violates condition 2 easily.

Although it is straightforward to compare var$[\hat{\mu}_{HT}]$ (3) and var$[\hat{\mu}]$, it cannot be easily interpreted since var$[\hat{\mu}_{HT}]$ involves the intersection probabilities. However, empirical results in Section 7 show that $\hat{\mu}_{HT}$ is consistently more efficient than $\hat{\mu}$.

7 Experiments and Discussion

This section presents quantitative evaluations of the ACS scheme and its estimators for wide area prospecting with a team of four robots; this is the number of robots that will be fielded on NASA Ames Research Center's Moonscape for our project's real-world test and evaluation. The experiments were performed using Webots, a mobile robot simulator (http://www.cyberbotics.com). 16 directed distance sensors with 0.3 m range were modelled around its body of 0.32 m (L) $\times\, 0.27$ m (W) $\times\, 0.2$ m (H). Each robot could sense its global position through GPS[1], and communicate spectrometric and tour data with the base. The robots used the potential fields method [9] for navigation between sampling units and obstacle avoidance. Each robot could move at a maximum speed of 0.425 m/s and consumed about 28.2 J/m. It used the Alpha Particle X-Ray Spectrometer (APXS) [15] (1.3 W) for sampling, which required about 2 hours to obtain a high-quality x-ray spectrum of the mineral content. So, sampling each unit would use about 9.5 kJ. The 6.46 km $\times$ 4.61 km prospecting region is discretized into a 28×20 grid of sampling units such that each unit's width is about 231 m. The robots were placed at a sampling unit in the center of the region and had to rendezvous at this same unit after all selected units were sampled.

To compare the performance of the estimators, the RMSE criterion is used to measure their quality:

$$\text{RMSE}[t] = \left[\frac{1}{R} \sum_{i=1}^{R} (\tau_i - \mu)^2 \right]^{\frac{1}{2}}$$

[1]Deployment of space exploration infrastructure would ultimately result in GPS or similar localization capability on the Moon and Mars. If this is absent, the current technique of a sun-seeking sensor combined with local inertial navigation can be used.

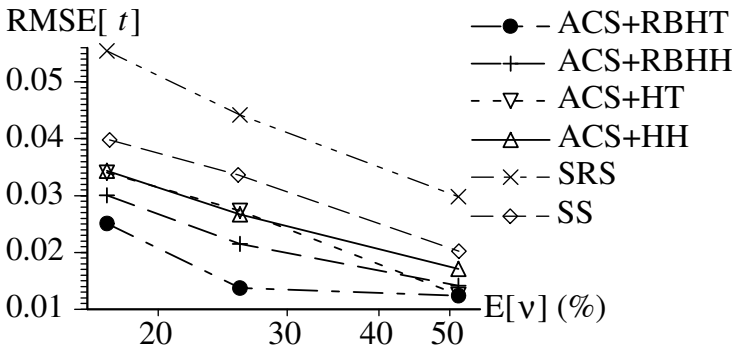
where $R = 20$ is the number of test runs, τ_i is the mean mineral content estimate obtained in test run i.

Using this measure, a quantitative test was conducted to compare the estimators described above. For ACS, the initial sample size n_1 was 40, 80 or 240 sampling units. After 20 test runs for each n_1, it resulted in an average final sample size E[ν] of approximately 95, 145, and 288 units, which corresponded to 17.1%, 25.7%, and 51.4% of the 560 total sampling units. The SRS and SS schemes were conducted using the same sample sizes as E[ν].

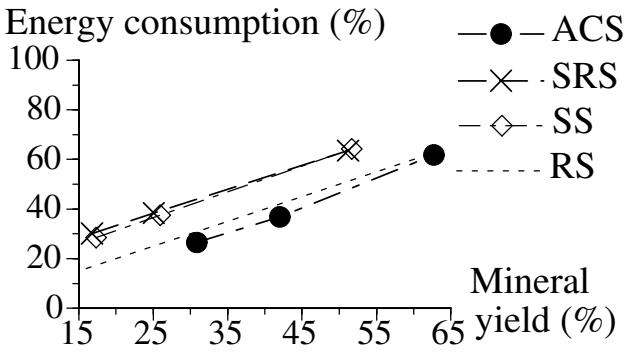
Test results (Fig. 3a) show that the ACS estimators perform better than the non-ACS estimators. Among the ACS estimators, the Rao-Blackwellized estimators achieved lower RMSE. In particular, the differences in performance between the estimators were the most significant at the sample size of 145 units. This implies that the ACS estimators, especially the Rao-Blackwellized ones, are practically more appealing because more accurate mineral density estimates can be obtained with a reasonably small sample size. If the sample size decreases too much below 17.1%, the performance of all estimators will converge since their behavior will be more and more alike. If the sample size increases too much beyond 51.4%, their performance will also converge due to increasing similarity in information from the prospecting region. Using t-tests ($\alpha = 0.1$), the differences in RMSEs between the estimators have been verified to be statistically significant if these differences are more than 0.006 for the sample sizes of 95 and 145 units, and more than 0.004 for the sample size of 288 units (see Appendix D). Note that the biased sample mean estimator $\bar{\mu}$ under the ACS scheme is not included in Fig. 3a; it has extremely large RMSEs of 0.525, 0.406, and 0.143 corresponding to 17.1%, 25.7%, and 51.4% of the total sampling units.

To compare the system performance of the sampling schemes, the optimality criteria mentioned in Section 2 are considered: minimizing (1) total energy consumption of all robots, and (2) maximum mission time of any robot. Fig. 3b and c show the results after 20 test runs for the first and second criterion respectively; the mineral yield, energy consumption, and mission time recorded for the various sampling strategies are given as a percentage of the corresponding values for RS (i.e., complete sampling of 560 units). Note that each strategy (other than RS) has three different records in its plot, which correspond to E[ν] of 95, 145, and 288 units; a smaller sample size gives a smaller mineral yield. The line for RS shows a constant ratio of energy consumption or mission time to mineral yield. We observe that the ACS strategy yields more minerals than SRS and SS with less energy and mission time. The differences in mineral yield, energy consumption or mission time between ACS and the other two strategies have been verified using t-tests ($\alpha = 0.1$) to be statistically significant except for that of energy consumption between ACS and SS with a sample size of 145 units.

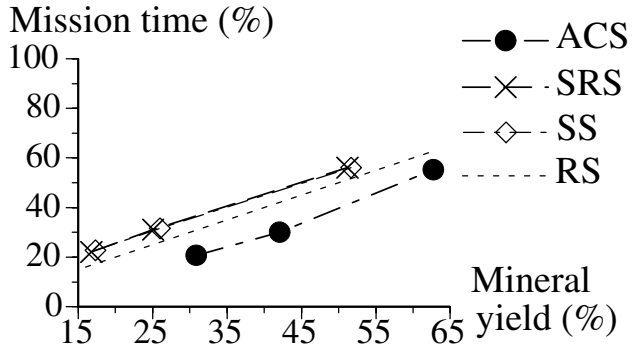
Furthermore, in contrast to SRS and SS, we observe that ACS falls below the dotted line of RS, which implies it achieves a lower ratio of energy consumption or mission time to mineral yield than RS. Hence, it is both energy and time efficient to utilize ACS for prospecting in place of RS. We also expect the advantage of ACS to increase when the sampling cost increases. For example, the alpha mode of APXS [15] requires at least 8 hours of sampling time, while the Mössbauer spectrometer [10] runs at 2 W and needs at least 6 hours. In our experiments for ACS, the spectrometry incurs 45% of the total energy consumption and 83% of the overall mission time for a typical sample size



(a)



(b)



(c)

Figure 3: Comparison of (a) RMSEs of different estimators and sampling strategies, (b) energy consumption of different sampling strategies, and (c) mission time of different sampling schemes.

of 288 units. These figures will increase substantially if the Mössbauer spectrometer is used instead.

8 Conclusion and Future Work

This paper describes the derivation of new low-error estimators within the ACS scheme and its application to multi-robot wide area prospecting. Quantitative experimental results in the prospecting task simulation have shown that the ACS scheme can yield more minerals with less resources and the Rao-Blackwellized ACS estimators can provide more precise mineral density estimates than previous methods. For our future work, we will apply these techniques on a larger robot team and real robots. Our planner will be improved using other minimum spanning tree heuristics or stochastic search strategies to reduce the tour lengths so that ACS can be even more efficient than RS. We will also consider the effect of noisy and multivariate mineral content data on our scheme and estimators. Lastly, adaptive systematic sampling will be examined.

References

[1] M. A. Batalin, M. Rahimi, Y. Yu, D. Liu, A. Kansal, G. S. Sukhatme, W. J. Kaiser, M. Hansen, G. J. Pottie, M. Srivastava, and D. Estrin. Call and response: Experiments in sampling the environment. In *Proc. ACM SenSys'04*, pages 25–38, 2004.

[2] D. Blackwell. Conditional expectation and unbiased sequential estimation. *Ann. Math. Stat.*, 18(1):105–110, 1947.

[3] C. A. Carlson. Spatial distribution of ore deposits. *Geology*, 19(2):111–114, 1991.

[4] N. A. C. Cressie. *Statistics for Spatial Data*. Wiley, NY, 2nd edition, 1993.

[5] G. N. Frederickson, M. S. Hecht, and C. E. Kim. Approximation algorithms for some routing problems. *SIAM J. Comput.*, 7(2):178–193, 1978.

[6] B. Glass and G. Briggs. Evaluation of human vs. teleoperated robotic performance in field geology tasks at a Mars analog site. In *Proc. 7th i-SAIRAS-03*, 2003.

[7] M. M. Hansen and W. N. Hurwitz. On the theory of sampling from finite populations. *Ann. Math. Stat.*, 14(4):333–362, 1943.

[8] D. G. Horvitz and D. J. Thompson. A generalization of sampling without replacement. *J. Am. Stat. Assoc.*, 47(260):663–685, 1952.

[9] O. Khatib. Real-time obstacle avoidance for manipulators and mobile robots. *Int. J. Robot. Res.*, 5(1):90–98, 1986.

[10] G. Klingelhöfer, R. V. Morris, B. Bernhardt, D. Rodionov, P. A. de Souza Jr., S. W. Squyres, J. Foh1, E. Kankeleit, U. Bonnes, R. Gellert, C. Schröder, S. Linkin, E. Evlanov, B. Zubkov, and O. Prilutski. Athena MIMOS II Mössbauer spectrometer investigation. *J. Geophys. Res.*, 108(E12), 2003.

[11] M. G. Lagoudakis, E. Markakis, D. Kempe, P. Keskinocak, A. Kleywegt, S. Koenig, C. Tovey, A. Meyerson, and S. Jain. Auction-based multi-robot routing. In *Proc. Robotics: Science and Systems*, 2005.

[12] P. G. Lucey, D. T. Blewett, and B. L. Jolliff. Lunar iron and titanium abundance algorithms based on final processing of Clementine ultraviolet-visible images. *J. Geophys. Res.*, 105(E8):20,297–20,305, 2000.

[13] P. G. Lucey, G. J. Taylor, and E. Malaret. Abundance and distribution of iron on the moon. *Science*, 268(5214):1150–1153, 1995.

[14] R. McCartney and H. Sun. Sampling and estimation by multiple robots. In *Proc. 4th ICMAS-00*, pages 415–416, 2000.

[15] R. Rieder, R. Gellert, J. Brückner, G. Klingelhöfer, G. Dreibus, A. Yen, and S. W. Squyres. The new Athena alpha particle X-ray spectrometer for the Mars exploration rovers. *J. Geophys. Res.*, 108(E12):7,1–7,13, 2003.

[16] M. M. Salehi. Rao-Blackwell versions of the Horvitz-Thompson and Hansen-Hurwitz in adaptive cluster sampling. *Environ. Ecol. Stat.*, 6(2):183–195, 1999.

[17] G. A. F. Seber and S. K. Thompson. Environmental adaptive sampling. In G. P. Patil and C. R. Rao, editors, *Environmental Statistics*, volume 12 of *Handbook of Statistics*, pages 201–220. North-Holland/Elsevier Sci. Publ., NY, 1994.

[18] P. D. Spudis and G. J. Taylor. The roles of humans and robots as field geologists on the moon. In W. W. Mendell, editor, *2nd Conference on Lunar Bases and Space Activities of the 21st Century*, NASA Conf. Pub. 3166, pages 307–313. 1992.

[19] G. J. Taylor and L. M. V. Martel. Lunar prospecting. *Adv. Space Res.*, 31(11):2403–2412, 2003.

[20] S. K. Thompson. *Sampling*. John Wiley & Sons, Inc., NY, 2002.

A Proof of Theorem 1

(a) Since $\mathcal{D}$ is minimally sufficient for $\boldsymbol{\theta}$, $\mathrm{E}[t|\mathcal{D}]$ does not depend on $\boldsymbol{\theta}$. Hence $t_{\mathcal{D}}$ does not depend on $\boldsymbol{\theta}$ and can be regarded as an estimator;
(b) $\mathrm{E}[t_{\mathcal{D}}] = \mathrm{E}_{\mathcal{D}}\mathrm{E}[t|\mathcal{D}] = \mathrm{E}[t]$;

$$\begin{aligned}(c)\ \mathrm{MSE}[t] = &\ \mathrm{E}[(t-\mu)^2] = \mathrm{E}[(t-t_{\mathcal{D}}+t_{\mathcal{D}}-\mu)^2] \\ = &\ \mathrm{E}[(t-t_{\mathcal{D}})^2] + \mathrm{MSE}[t_{\mathcal{D}}]\end{aligned} \tag{26}$$

since

$$\begin{aligned}\mathrm{E}[(t-t_{\mathcal{D}})(t_{\mathcal{D}}-\mu)] = &\ \mathrm{E}_{\mathcal{D}}\mathrm{E}[(t-t_{\mathcal{D}})(t_{\mathcal{D}}-\mu)|\mathcal{D}] \\ = &\ \mathrm{E}_{\mathcal{D}}(t_{\mathcal{D}}-\mu)\mathrm{E}[(t-t_{\mathcal{D}})|\mathcal{D}] \\ = &\ \mathrm{E}_{\mathcal{D}}(t_{\mathcal{D}}-\mu)(\mathrm{E}[t|\mathcal{D}]-t_{\mathcal{D}}) \\ = &\ 0\ .\end{aligned} \tag{27}$$

It follows from (26) that $\mathrm{MSE}[t_{\mathcal{D}}] \leq \mathrm{MSE}[t]$ with equality if and only if $\mathrm{E}[(t-t_{\mathcal{D}})^2] = 0$ (i.e., $t = t_{\mathcal{D}}$ with probability 1). Strict inequality will occur if t is different from $t_{\mathcal{D}}$ over a set of nonzero probability.

B Proof of Corollary 1

Given that t is unbiased, $\mathrm{E}[t] = \mu$. Then

$$\begin{aligned} \mathrm{MSE}[t] &= \mathrm{E}[(t-\mu)^2] \\ &= \mathrm{E}[(t-\mathrm{E}[t]+\mathrm{E}[t]-\mu)^2] \\ &= \mathrm{var}[t] + (\mathrm{E}[t]-\mu)^2 \\ &= \mathrm{var}[t]\ . \end{aligned} \tag{28}$$

Since $t_{\mathcal{D}}$ is also unbiased (Theorem 1), $\mathrm{E}[t_{\mathcal{D}}] = \mu$. Corollary 1 follows from (26) by using (28) to replace $\mathrm{MSE}[t]$ and $\mathrm{MSE}[t_{\mathcal{D}}]$ with $\mathrm{var}[t]$ and $\mathrm{var}[t_{\mathcal{D}}]$ respectively.

C Derivation: var$[\hat{\mu}_{HT} \mid \mathcal{D}]$

$\hat{\mu}_{HT}$ (2) for the gth combination can be expressed as

$$\hat{\mu}^g_{HT} = \sum_{i\in\mathcal{E}} \frac{y^*_i}{N\pi_i} I_{gi} = \sum_{i\in\mathcal{F}_1\cup\mathcal{F}_3} \frac{y^*_i}{N\pi_i} + \sum_{i\in\mathcal{F}_2} \frac{y^*_i}{n_1} I_{gi} \tag{29}$$

since $\pi_i = n_1/N$ for networks of size one. Note that $I^2_{gi} = I_{gi}$, $\sum_{g=1}^{\xi} I_{gi} = \xi_1$ for $i \in \mathcal{F}_2$, $I_{gi}I_{gj} = 1$ if the gth combination contains at least one unit each from $\mathcal{B}_i$ and $\mathcal{B}_j$, and $I_{gi}I_{gj} = 0$ otherwise. Since networks i and j in $\mathcal{F}_2$ are each of size 1, $\sum_{g=1}^{\xi} I_{gi}I_{gj} = \xi_2$, i.e., the number of combinations containing any two edge units in $\mathcal{F}_2$. By substituting (14) and (29) into the second term of (10),

$$\begin{aligned} \mathrm{var}[\hat{\mu}_{HT}|\mathcal{D}] &= \frac{1}{\xi}\sum_g (\hat{\mu}^g_{HT} - \hat{\mu}_{RBHT})^2 \\ &= \frac{1}{\xi}\sum_{g=1}^{\xi}\left(\sum_{i\in\mathcal{F}_2}\frac{y^*_i}{n_1}I_{gi} - \frac{\xi_1}{\xi}\sum_{i\in\mathcal{F}_2}\frac{y^*_i}{n_1}\right)^2 \\ &= \frac{1}{n_1^2\xi^3}\sum_{g=1}^{\xi}\left(\sum_{i\in\mathcal{F}_2} y^*_i(\xi I_{gi} - \xi_1)\right)^2 \\ &= \frac{1}{n_1^2\xi^3}\sum_{g=1}^{\xi}\sum_{i\in\mathcal{F}_2} y^{*2}_i(\xi I_{gi} - \xi_1)^2 \\ &\quad + \frac{2}{n_1^2\xi^3}\sum_{g=1}^{\xi}\sum_{i\in\mathcal{F}_2}\sum_{j<i} y^*_i y^*_j(\xi I_{gi} - \xi_1)(\xi I_{gj} - \xi_1) \\ &= \frac{1}{n_1^2\xi^3}\sum_{i\in\mathcal{F}_2} y^{*2}_i \sum_{g=1}^{\xi}(\xi I_{gi} - \xi_1)^2 \\ &\quad + \frac{2}{n_1^2\xi^3}\sum_{i\in\mathcal{F}_2}\sum_{j<i} y^*_i y^*_j \sum_{g=1}^{\xi}(\xi I_{gi} - \xi_1)(\xi I_{gj} - \xi_1) \end{aligned}$$

$$
\begin{aligned}
= \quad & \frac{1}{n_1^2\xi^3}\sum_{i\in\mathcal{F}_2} y_i^{*2}\sum_{g=1}^{\xi}(\xi_1^2+\xi^2 I_{gi}^2-2\xi\xi_1 I_{gi}) \\
& +\frac{2}{n_1^2\xi^3}\sum_{i\in\mathcal{F}_2}\sum_{j<i} y_i^* y_j^*\sum_{g=1}^{\xi}(\xi^2 I_{gi}I_{gj}-\xi\xi_1 I_{gi}-\xi\xi_1 I_{gj}+\xi_1^2) \\
= \quad & \frac{1}{n_1^2\xi^3}\sum_{i\in\mathcal{F}_2} y_i^{*2}\left(\sum_{g=1}^{\xi}\xi_1^2+\xi^2\sum_{g=1}^{\xi} I_{gi}-2\xi\xi_1\sum_{g=1}^{\xi} I_{gi}\right) \\
& +\frac{2}{n_1^2\xi^3}\sum_{i\in\mathcal{F}_2}\sum_{j<i} y_i^* y_j^*\left(\xi^2\sum_{g=1}^{\xi} I_{gi}I_{gj}\right. \\
& \left.-\xi\xi_1\sum_{g=1}^{\xi} I_{gi}-\xi\xi_1\sum_{g=1}^{\xi} I_{gj}+\sum_{g=1}^{\xi}\xi_1^2\right) \\
= \quad & \frac{1}{(n_1\xi)^2}\sum_{i\in\mathcal{F}_2} y_i^{*2}(\xi_1\xi-\xi_1^2) \\
& +\frac{2}{(n_1\xi)^2}\sum_{i\in\mathcal{F}_2}\sum_{j<i} y_i^* y_j^*(\xi_2\xi-\xi_1)^2 \\
= \quad & \frac{1}{(n_1\xi)^2}\left((\xi_1\xi-\xi_1^2)\sum_{i\in\mathcal{F}_2} y_i^{*2}\right. \\
& \left.+2(\xi_2\xi-\xi_1)^2\sum_{i\in\mathcal{F}_2}\sum_{j<i} y_i^* y_j^*\right).
\end{aligned}
$$

D Significance levels from t-tests on similarity in RMSEs between estimators

E[ν] = 95	SS	SRS	ACS+HH	ACS+HT	ACS+RBHH
ACS+RBHT	0.01	0.00	0.07	0.05	0.16
ACS+RBHH	0.05	0.00	0.25	0.24	
ACS+HT	0.16	0.00	0.48		
ACS+HH	0.20	0.00			
SRS	0.01				

E[ν] = 145	SS	SRS	ACS+HH	ACS+HT	ACS+RBHH
ACS+RBHT	0.00	0.00	0.02	0.02	0.01
ACS+RBHH	0.00	0.01	0.17	0.16	
ACS+HT	0.10	0.04	0.46		
ACS+HH	0.07	0.03			
SRS	0.10				

E[ν] = 288	SS	SRS	ACS+HH	ACS+HT	ACS+RBHH
ACS+RBHT	0.00	0.00	0.10	0.48	0.25
ACS+RBHH	0.01	0.00	0.20	0.25	
ACS+HT	0.00	0.01	0.10		
ACS+HH	0.16	0.01			
SRS	0.02				