

Performance of Shrinkage Estimators for Bioburden Density Calculations in Planetary Protection Probabilistic Risk Assessment

Andrei Gribok^a, Arman Seuylemezian^b

^aIdaho National Laboratory, Idaho Falls, USA

^bCalifornia Institute of Technology, Jet Propulsion Laboratory, Pasadena, USA

Abstract: Planetary protection (PP) is a discipline that focuses on minimizing the biological contamination of spacecraft to ensure compliance with international policy. The National Aeronautics and Space Administration has developed a set of requirements (NPR 8715.24) based on recommendations from the Committee on Space Research that each mission must comply with regarding both forward and backward PP. Biological cleanliness requirements to target bodies, such as Mars, include spacecraft assembly control and the direct testing of the microbial bioburden of different components to comply with PP requirements. The data for each component are collected using either swabs or wipes. For each component, a number of samples are collected on one given date or on several different dates along the course of the part assembly. Given the clean spacecraft, on the average 93% of the swabs and 63% of the wipes have no colony forming units (CFU) count at 72 hours, resulting in ~85% of the 39,379 petri dishes yielding 0 CFU. Due to low CFU counts and small sampling areas, given the Poisson distributional model, the bioburden density estimates have inflated variance and confidence intervals. Shrinkage estimators are standard tools to deal with large variance and estimate inconsistencies. This paper presents the performance results of six shrinkage estimators along with the maximum likelihood, population-average, and zero estimators applied to the bioburden density estimation using InSight mission data. The results show that, for poolable data sets, the best estimator is population average, while for nonpoolable data sets, the Tsui estimator along with the empirical Bayes estimator produced the lowest mean squared error.

1. INTRODUCTION

The planetary protection (PP) discipline's primary objective is to minimize the inadvertent microbial contamination of other planetary bodies via hitchhiking microbes present on robotic spacecraft destined for these planetary bodies. PP engineers thereby constantly monitor, assess, and mitigate the microbiome of spacecraft surfaces and cleanroom assembly environments to ensure the responsible exploration of the solar system. Although a suite of molecular techniques have been used to thoroughly characterize and profile the microbiome of various cleanroom environments and spacecraft, the gold standard remains the physical enumeration of microbes via culturing of samples directly taken from spacecraft and associated surfaces. These samples are then subjected to laboratory processing and result in colony forming units (CFU) counts that are ultimately represented as bioburden density (CFU/m²). However, due to technical, budgetary, and programmatic constraints, only a manageable portion of the entire spacecraft surface is directly sampled. PP engineers are then tasked with estimating the bioburden density and total microbial burden of the entire spacecraft in order to demonstrate compliance with stringent requirements set forth by the National Aeronautics and Space Administration (NASA), which are specific to each mission and its unique scientific objectives. The InSight mission, to explore the interior structure and processes of Mars with two primary instruments, had an at-launch bioburden requirement for the entire spacecraft of 1.50×10^5 spores while the cruise stage had a requirement of 5×10^5 spores. The landed spacecraft hardware had to remain $<3 \times 10^5$ spores while maintaining a bioburden density of <300 spores/m².

To manage and track the bioburden density and total bioburden throughout the lifecycle of the entire mission, PP engineers maintain a PP equipment list, which tracks the surface area, related bioburden, and overall assembly hierarchy of each spacecraft component, subsystem, and system. Historically, several

different statistical approaches to estimate bioburden density were developed and adopted to the needs of each unique spacecraft mission highlighted in References [1],[2],[3]. To meet the demands of more biologically sensitive spacecraft missions and develop a robust approach to account for uncertainties associated with the sampling and recovery process, a Bayesian statistical model was developed and applied to perform bioburden calculations on datasets generated from spacecraft missions [4].

This study compares and contrasts different shrinkage estimators and their performance in evaluating microbial bioburden data collected from the InSight mission.

2. DATA COLLECTION AND PREPROCESSING

Data samples were collected using either cotton Puritan (Guilford, ME) 806C swabs or TexWipe (Kernersville, NC) TX3211 polyester wipes. Swabs sampled a maximum of a 0.0025 m² surface area while wipes sampled up to a 1 m² surface area. Prior to sampling, both swabs and wipes were premoistened with sterile water to improve microbial capture. After sampling, swabs were resuspended in 10 ml of distilled water. While in storage, TX3211 wipes were stored in sterile glass screw cap bottles prior to transporting them for processing. To begin the wet laboratory processing, wipes were suspended in 200 ml of polypropylene rinse solution. Both swab and wipe samples were then subjected to 2 minutes of sonication and 15 minutes of an 80°C heat shock. The heat shock step selects for the hardy subpopulation of microorganisms that represent the higher threshold of heat tolerance. These organisms are a proxy for the total microbial bioburden present and present the highest risk of organisms surviving the cleaning procedures (primarily dry heat exposure) and potential for surviving space conditions. A 2 ml aliquot from either the swab or wipe samples (suspended in 10 ml and 200 ml of suspension respectively) are taken and plated onto 4 and 25 petri dishes respectively, resulting in a total of 8 ml’s being plated from the 10 ml swab samples and 50 ml’s being plated from the 200 ml wipe sample. Due to this experimental procedure, the swabs assume a pour fraction of 0.8 and the wipes 0.25, representing the portion of the total sample solution plated and analyzed for CFU counts. These plates were then overlaid with ~20 ml of tryptic soy agar nutrient media to promote CFU growth. Once plates dried, they were incubated at 32°C for 72 hours, and CFUs were counted at 24-hour intervals for 72 hours. The surface area sampled are estimated by the samplers while appropriate negative and positive controls ensure aseptic processing.

For estimator analyses, the raw data for each component were represented by pairs (x_i, E_i) , $i=1,2,\dots,N$, where x_i is the CFU counts for the i -th swab or wipe sample and E_i is the exposure calculated as the area covered with a swab or wipe multiplied by the corresponding pour fraction and N is the number of samples collected for a component. We have used “exposure” and “effective sampled area” in this paper interchangeably. This study analyzed eight different InSight components that represent a range of different total CFU counts, effective areas sampled, and total areas of the components. The components were selected to represent typical ranges of CFU counts, exposures, and percent of sampled area. Table 1 summarizes this data.

Table 1: Summary of Bioburden Data for the Eight Components.

Component	CFU Count	Number of Samples	Area Sampled, m ²	Total Exposure: Area Sampled*Pour Fraction, m ²	Total Surface of the Component, m ²	% Sampled=Area Sampled/Total Area
9	0	49	0.6031	0.2167	0.7580	79.5650
38	12	27	3.1050	0.8065	10.0000	31.0500
67	0	27	2.4200	0.6160	2.7400	88.3212
169	1	17	0.2400	0.1920	0.5850	41.0260
243	5	34	0.2800	0.1140	0.2980	93.9600
261	52	24	0.0600	0.0480	0.3120	19.2310
283	5	8	4.5710	1.1427	12.0000	38.0920
300	1	9	2.6600	0.6705	5.0000	53.2000

3. DATA POOLABILITY AND RISK FUNCTION

Prior to finding the bioburden density, a number of assumptions for the CFU counts need to be made:

1. The probability of finding a CFU on any specified small exposure area is proportional to the exposure area and does not depend on where that exposure area is located. In other words, the bioburden density does not depend on location.
2. The probability of finding more than one CFU on a given small exposure area is negligible in comparison with probability of finding exactly one CFU on that area.
3. Finding CFUs on disjoint exposure areas is a statistically independent event.

If the above assumptions hold, the probabilistic model applied to the number of CFUs counts is a Poisson distribution with the probability mass function:

$$P(X = x|\lambda) = \frac{(\lambda \cdot E)^x}{x!} \cdot e^{-\lambda \cdot E}, \lambda \geq 0, x = 0, 1, \dots \quad (1)$$

where X is the random variable describing CFU counts, x is the actual number of CFUs found on the exposure area E , and λ is the bioburden density or expected number of CFUs per unit of exposure, which is unknown and the subject of the statistical inference. If the observed CFU count is x_i for a given exposure E_i , λ_i can be estimated as:

$$\hat{\lambda}_i = \frac{x_i}{E_i}, i = 1, \dots, N \quad (2)$$

where N is the number of samples.

This estimate is the maximum likelihood estimate (MLE) [1] currently used by NASA to evaluate the bioburden density and total CFU counts for biologically sensitive missions [1]. The MLE allows the bioburden density for each sample to be examined separately, and it has a number of desirable statistical properties [5]. However, it also has a number of shortcomings, such as large variance, and most importantly, for small number of observed CFUs, it can overfit the data. For example, if the number of CFUs registered on an exposure surface is zero, MLE will produce a bioburden density estimate of zero, which is highly unlikely, as achieving absolute cleanliness is practically impossible considering the presence of humans during spacecraft assembly. On the other hand, if Assumption 1 is valid and each sample has the same λ , the data from different samples can be pooled to obtain a single estimate of λ , which can be used for each sample. In this case, the pooled λ is estimated as:

$$\hat{\lambda} = \frac{\sum_{i=1}^N x_i}{\sum_{i=1}^N E_i} \quad (3)$$

If Assumption 1 is correct, the data can be pooled for a given component to obtain a single estimate of bioburden density, which can be used as an estimate for each sample in an assay. The validity of assumption 1 can be verified by a number of techniques [6], such as graphical techniques or more rigorous statistical tests. The graphical technique, uses a caterpillar plot for λ_i , estimated separately for each sample. On the plot, each λ_i is accompanied by its confidence interval. The analysis of the plot can reveal trends, outliers, and differences in bioburden densities for different samples.

If the graphical analysis indicates a significant λ variability among the samples, it suggests that a more rigorous statistical test is in order to quantify the difference between samples. The statistical test normally employed in this case is the “chi-square” test [6]. The chi-square test assumes that H_0 is the hypothesis of

the same λ for all samples, while the alternative hypothesis H_1 is that the λ s are different for different samples. If H_0 is true, the data from different samples can be pooled and the same population-average λ can be used for all samples. On the other hand, if H_0 is rejected, each sample needs to be treated separately, that is as having sample-specific λ s. To perform the chi-squared test, a χ^2 statistic is:

$$\chi^2 = \sum_{i=1}^N \frac{(x_i - ex_i)^2}{ex_i} \quad (4)$$

where x_i is the observed CFU count, ex_i is the expected CFU count for i -th sample calculated as $ex_i = \hat{\lambda} \cdot E_i$ with $\hat{\lambda}$ being the pooled population-average bioburden density as defined in Eq. 3, and E_i is the exposure for the i -th sample. The statistic has an χ^2 distribution with $N-1$ degrees of freedom, and its $1-\alpha$ quantile is compared with the value of the statistic calculated using Eq. 4. If the calculated value of the statistics is larger than the $1-\alpha$ quantile, the null hypothesis is rejected at the α significance level. For example, the Component 261 data in Table 1 has the statistics value of 181.5 while the 95% quantile is 35.1, indicating that the null hypothesis of equal λ s for all samples has to be rejected at the 5% significance level. The rejection of the null hypothesis calls for utilizing sample-specific λ s for each sample. However, using single sample data with small exposure for each sample increases the uncertainty of bioburden density estimates. Pooling the data from different samples would increase the total exposure, as evident from Eq. 3; however, the statistical test does not support the evidence of the same λ for each sample. Fortunately, statistical theory offers a compromise between the two alternatives, which can be derived in the Bayesian inference framework [7],[8]. The Bayesian analysis requires assuming the Gamma-Poisson compound distribution model shown in Eq. 5:

$$\underbrace{P(\lambda_{post}/x)}_{\text{Posterior}} = \underbrace{\mathcal{G}(x + \alpha_{prior}, E + \beta_{prior})}_{\alpha_{post}} = \frac{\underbrace{\frac{(\lambda \cdot E)^x}{x!} e^{-\lambda \cdot E}}_{\text{Likelihood}} \cdot \underbrace{\frac{\beta_{prior}^{\alpha_{prior}} \cdot \lambda^{\alpha_{prior}-1} \cdot e^{-\lambda \cdot \beta_{prior}}}{\Gamma(\alpha_{prior})}}_{\text{Prior}}}{\int_0^{\infty} \underbrace{\frac{(\lambda \cdot E)^x}{x!} e^{-\lambda \cdot E}}_{\text{Likelihood}} \cdot \underbrace{\frac{\beta_{prior}^{\alpha_{prior}} \cdot \lambda^{\alpha_{prior}-1} \cdot e^{-\lambda \cdot \beta_{prior}}}{\Gamma(\alpha_{prior})}}_{\text{Prior}} d\lambda}_{\text{Evidence}}} \quad (5)$$

where λ is the bioburden density of a sample, λ_{post} is the posterior value of λ , α_{prior} and β_{prior} are prior values of prior gamma distribution parameters, Γ is the gamma function, e is the sample's exposure value, x is the sample's CFU count, \mathcal{G} is the gamma distribution function, and α_{post} and β_{post} are posterior parameters of the gamma distribution. The posterior distribution can be summarized by a number of point estimates, such as its mean value, maximum value, or median. Under the quadratic loss function, the point estimate that delivers the smallest loss is posterior mean value [9]. The analysis presented in this paper assumes the quadratic loss and posterior mean value as summary statistics of the posterior distribution. Quadratic loss for a single sample is defined as:

$$L(\lambda_{true}, \hat{\lambda}(x)) = (\lambda_{true} - \hat{\lambda}(x))^2 \quad (6)$$

where λ_{true} is the true unknown value of the bioburden density, called the estimand, and $\hat{\lambda}(x)$ is bioburden density estimate, which depends on data x . The risk of an estimator $\hat{\lambda}(x)$ is defined as mathematical expectation of the loss function taken over the data distribution, generally:

$$R(\lambda_{true}, \hat{\lambda}(x)) = \sum_x (\lambda_{true} - \hat{\lambda}(x))^2 \cdot p(x; \lambda_{true}) \quad (7)$$

This risk function is known as the mean squared error (MSE) and is used for all the results in this paper. The MSE is a function of the true unknown λ as well as the data.

4. MAXIMUM LIKELIHOOD AND SHRINKAGE ESTIMATORS

A natural estimator of λ is the MLE, which in case of a Poisson distribution in Eq. 1, takes the form of Eq. 8, where x is the number of observed CFUs and e is the exposure:

$$\hat{\lambda}_{MLE} = \frac{x}{E} \quad (8)$$

If only one sample of the data is available, this is the best estimate in the sense of minimizing the risk in Eq. 7. However, if several measurements are available and the problem is to estimate several λ s simultaneously, the MLE is no longer the best, and there are estimators with smaller risk [7],[8],[9],[10],[12],[13]. For several measurements, all λ s could be the same, thus making the data set poolable, or λ s may differ from sample to sample, making the data set nonpoolable. The problem of simultaneously estimating several λ s generated from a family of Poisson distributions can be formulated as follows: suppose we have N Poisson variables X_1, X_2, \dots, X_N with corresponding exposures E_1, E_2, \dots, E_N generated from Poisson distributions with $\lambda_1, \lambda_2, \dots, \lambda_N$ and we wish to obtain the vector of λ estimates $\widehat{\lambda}_1, \widehat{\lambda}_2, \dots, \widehat{\lambda}_N$, using the entire vector of observations X_1, X_2, \dots, X_N . The usual MLE estimator in Eqs. 2 and 8 is an unbiased minimum variance estimator, which is admissible for $N=1$ [11],[10]. Two different options concerning λ s can be considered for the above problem statement, the first option is when all λ s are the same and the data can be pooled to obtain an estimate presented in Eq. 3. The second option is when all λ s are different and each measurement needs to be treated separately. In both cases, our goal is to obtain an estimate that minimizes the risk function in Eq. 7. While these two options represent extreme cases, there is a continuous range of possibilities between them represented by shrinkage estimators that perform partial pooling by relying on the current and other measurements or some prior information. The idea of shrinkage estimators is best presented using the Bayesian framework, specifically in our case the gamma-Poisson compound distribution model shown in Eq. 5. Under this model for a single measurement, the mean value of posterior distribution can be written as:

$$E(\lambda|x) = \frac{x + \alpha_{prior}}{e + \beta_{prior}} = \left[\frac{E}{E + \beta_{prior}} \right] \cdot \left(\frac{x}{E} \right) + \left[\frac{\beta_{prior}}{E + \beta_{prior}} \right] \cdot \left(\frac{\alpha_{prior}}{\beta_{prior}} \right) = [1 - B] \cdot \left(\frac{x}{E} \right) + [B] \cdot \left(\frac{\alpha_{prior}}{\beta_{prior}} \right) = \left(\frac{x}{E} \right) - [B] \cdot \left(\frac{x}{E} - \frac{\alpha_{prior}}{\beta_{prior}} \right) = \left(\frac{x}{E} \right) - \left[\frac{\beta_{prior}}{E + \beta_{prior}} \right] \cdot \left(\frac{x}{E} - \frac{\alpha_{prior}}{\beta_{prior}} \right) \quad (9)$$

where

$$B = \frac{\beta_{prior}}{E + \beta_{prior}} \leq 1 \quad (10)$$

is the shrinkage factor, α_{prior} and β_{prior} are parameters of the prior gamma distribution. An analysis of Eq. 9 reveals that the conjugate gamma-Poisson model performs shrinkage estimates by subtracting a fraction, defined by B , of the difference between MLE and prior mean expressed as $\frac{\alpha_{prior}}{\beta_{prior}}$ from the MLE estimate thus it is shrunk towards the prior mean. For $B=1$, the gamma-Poisson model is equivalent to replacing the MLE estimate with a prior mean, while for $B=0$, it is equivalent to the MLE. If the prior distribution variance is zero (β_{prior} is going to infinity, meaning data are poolable), every measurement is reset to the population average (prior average)—accomplishing complete pooling. If the prior distribution variance is infinity (β_{prior}

is going to zero), every measurement is left untouched—resulting in no pooling at all. Between these two extremes, the gamma-Poisson model performs partial pooling. If exposure is zero, all measurements are set to the prior mean because the collected data are sampled from an infinitesimal surface. If exposure is infinite, the measurement is set to the MLE, as we have an infinite amount of data and prior distribution does not matter anymore. Using Eq. 9, we can express a general shrinkage estimator:

$$\hat{\lambda}_i = \frac{x_i}{E_i} - B \cdot \left(\frac{x_i}{E_i} - T \right), i = 1, \dots, N \quad (11)$$

where T is the shrinkage target value and B is the shrinkage factor. Setting T=0, we obtained an estimator that shrank all MLE estimates toward zero:

$$\hat{\lambda}_i = \frac{x_i}{E_i} - B \cdot \left(\frac{x_i}{E_i} \right), i = 1, \dots, N \quad (12)$$

The performance of shrinkage estimators critically depends on the T and B value selection, and the main difference between shrinkage estimators is how those values are inferred. For the gamma-Poisson model, the method of moments (MOM) [1] can be applied to infer the parameters of prior gamma distributions, α_{prior} and β_{prior} . Having inferred the parameters, we applied Eq. 9 to construct shrinkage estimators that we call Empirical Bayes with MOM parameter inference (EB-MOM). In addition to the EB-MOM estimator, there are a number of different shrinkage estimators for the simultaneous estimation of Poisson means. One of the most popular is the Clevenson-Zidek (CZ) estimator [10]:

$$\hat{\lambda}_i = \frac{x_i}{E_i} - \frac{\gamma + N - 1}{\sum_{E_i} \frac{x_i}{E_i} + \gamma + N - 1} \cdot \left(\frac{x_i}{E_i} \right), i = 1, \dots, N \quad (13),$$

which shrinks the MLE estimate toward zero with parameter $0 \leq \gamma \leq N - 1$. Notice, for large values of $\frac{x_i}{E_i}$, the estimator is close to the MLE. Peng [12] has proposed another shrinkage estimator:

$$\hat{\lambda}_i = \frac{x_i}{E_i} - \frac{(N - N_0 - 2)_+}{S_h} \cdot h \left(\frac{x_i}{E_i} \right), i = 1, \dots, N \quad (14)$$

where N_0 is the number of observations equal to zero, $(N - N_0 - 2)_+$ = maximum of 0 and $(N - N_0 - 2)_+$, $h \left(\frac{x_i}{E_i} \right) = \sum_{j=1}^{\frac{x_i}{E_i}} \frac{1}{j}$, if $x_i=1, 2, \dots$; $h(0)=0, S_h = \sum_{j=1}^N h^2 \left(\frac{x_j}{E_j} \right)$. The Peng estimator also shrinks the MLE estimate towards zero. Zero-shrinking estimators usually perform very well for very small values of the true λ , as the shrinkage target in this case is close to the true value of the parameter. The original Peng estimator sets samples with zero CFUs to zero. In practice, nonzero estimates are desirable even though no CFUs are observed. In this case, a modified Peng estimator [12] can write the estimate of the true λ :

$$\hat{\lambda}_i = \frac{x_i}{E_i} - \frac{(N - N_0 - 2)_+}{S_h + N_0} \cdot h \left(\frac{x_i}{E_i} \right), i = 1, \dots, N \quad (15)$$

if $x_i=1, 2, \dots$, or minimum $\left\{ \frac{(N - N_0 - 2)_+}{S_h + N_0}, 1 - \frac{(N - N_0 - 2)_+}{S_h + N_0} \right\}$, if $x_i=0$. Both Peng estimators shrink the estimate towards zero. However, in practice, shrinkage towards other target values, such as the mean value or median of the observations, is often desirable. Estimators proposed by Tsui accomplish this task using [13]:

$$\hat{\lambda}_i = \frac{x_i}{E_i} - \frac{r(\pi)}{S_H} \cdot H \left(\frac{x_i}{E_i} \right), i = 1, \dots, N \quad (16)$$

where $H \left(\frac{x_i}{E_i} \right) = h \left(\frac{x_i}{E_i} \right) - h(\pi)$, $S_H = \sum_{j=1}^N H^2 \left(\frac{x_j}{E_j} \right)$, $r(\pi) = (N - \sum_{j=0}^{\pi} N_j - 3)_+$, and N_j is the number of observations equal to j. The function h is the same as in Peng's estimator in Eq. 14. The above estimator

shrinks the MLE towards the prespecified mean value of π . The final estimator we considered in this paper is the Tsui median estimator [13], which shrinks the MLE estimator towards the median of the data. If the median of the data is M , the Tsui median estimator is defined as:

$$\begin{aligned}
 H_i\left(\frac{x}{E}\right) &= 1 + \sum_{j=2}^{\frac{x_i-M}{E_i}} \frac{1}{(j+M)} \text{ if } \frac{x_i}{E_i} \geq M + 2 \\
 &= 1 \text{ if } \frac{x_i}{E_i} = M + 1 \\
 &= 0 \text{ if } \frac{x_i}{E_i} = M \\
 &= -b \text{ if } \frac{x_i}{E_i} \leq M
 \end{aligned} \tag{17}$$

where b is any positive constant. Larger b values can indicate a belief that λ s are nonzero. Further, $S_M = \sum_{j=1}^N H^2\left(\frac{x_j}{E_j}\right)$ is defined similar to the previous Tsui estimator and $r_M = (\text{the number of observations greater than } M - 2, 0)_+$. The median Tsui estimator is given by:

$$\hat{\lambda}_i = \frac{x_i}{E_i} - \frac{r_M}{S_M} \cdot H\left(\frac{x_i}{E_i}\right), i = 1, \dots, N \tag{18}$$

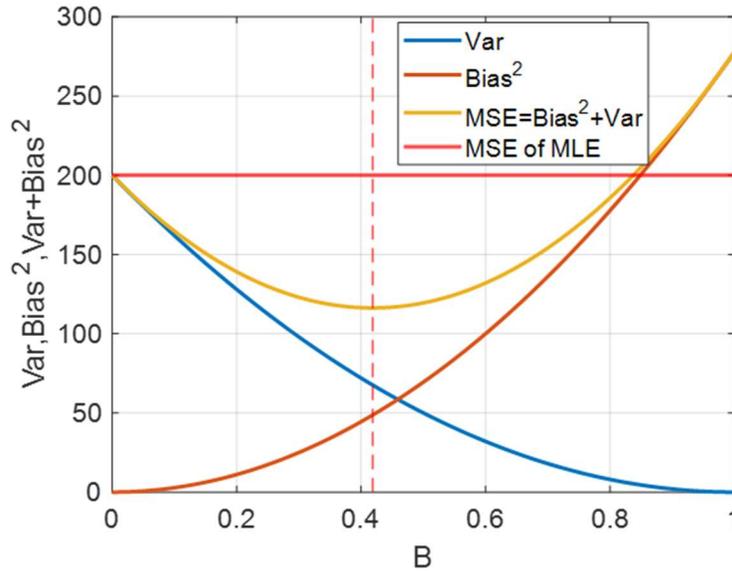
Despite the seemingly diverse nature of the proposed estimators, all shrinkage estimators exploit the same idea of trading a small bias for a significant reduction in estimators' variance. The MSE of the MLE estimator for a single measurement can be written as $\frac{\lambda_{true}}{e}$, so for small exposures, the loss can be rather large despite the fact that MLE is unbiased. For the generalized shrinkage estimator presented in Eq. 11, the MSE bias variance can be written as:

$$MSE_{Shrinkage} = \frac{\lambda_{true}}{e} \cdot (1 - B)^2 + (B \cdot (\lambda_{true} - T))^2 \tag{19}$$

where $0 \leq B \leq 1$ is the shrinkage factor and T is the shrinkage target. The first term on the right-hand side of Eq. 19 is the variance of the generalized shrinkage estimator, while the second term is the bias. For $B=0$, the MSE of the shrinkage estimator is equal to the MSE of the MLE estimator with no bias, while for $B=1$ the MSE of the shrinkage estimator is the square of the difference between the true value of λ and the preselected target value T . The values of B between zero and one are used to trade bias for variance. The bias-variance tradeoff as a function of B is illustrated in Figure 1.

Since the MLE MSE does not depend on B , it is a constant; as B increases, the bias of the estimator is also increasing while the variance is decreasing. The yellow curve in Fig.1 is Eq. 19 plotted as a function of B . Here, B has the optimal value where the MSE of the shrinkage estimator reaches the minimum, which is significantly lower than the MLE MSE. In this case, the optimal B value happened to be 0.4. More importantly, B has a wide range of values where the shrinkage estimator dominates the MLE (i.e., it has a lower MSE than the MLE). The closer the target value is to the true λ , the stronger the domination. Shrinkage parameter B controls the information exchange between prior information expressed as T and the data. If the target value is equal to the true λ , there is no bias, but the shrinkage estimator MSE is still lower than for the MLE. Knowing the true value of the parameter is the extreme case of prior information. Figure 1 can also be interpreted as the overfitting vs. underfitting tradeoff, since for small B values, the shrinkage estimator relies more on the data, and if the data is noisy, it is prone to overfitting. On the other hand, for large B values, the estimator tends to ignore data in favor of prior information, and thus, is prone to underfitting. The difference between shrinkage estimators presented in this paper is in how they select the shrinkage target value T and shrinkage factor B .

Figure 1: Bias-Variance Tradeoff for Shrinkage Estimators



To analyze the performance of different shrinkage estimators, we adopted the simulation approach as the true bioburden densities are unknown. From prior work on bioburden density estimation [1],[4],[2],[3], typical bioburden densities range from less than 1 CFU/m² up to 2,000 CFUs/m². To simulate having poolable data, for an equally spaced set of true λ s from this range and for swab and wipe exposures, we generated a vector of Poisson variables. Using this vector of Poisson variables and known exposure, we applied nine different estimators and plotted their MSE as a function of the true λ . The nine estimators were: MLE; zero estimator, which sets all estimates to zero regardless of the data; grand mean estimator, which pools all the data and is represented by Eq. 3; Empirical Bayes estimator, which uses the MOM to fit the prior distribution for λ ; CZ estimator; Peng; modified Peng estimators; and two Tsui estimators, one shrinking to a prespecified mean value and the other shrinking towards the data median.

To simulate nonpoolable data sets, we treated the equally spaced vector of true λ s as the vector of means of a gamma distribution with a variance of 10^5 , which is approximately the variance of the population variability for bioburden data. We also applied the same nine estimators mentioned above to nonpoolable data.

4. RESULTS AND DISCUSSION

The simulation results for poolable data sets for two exposures are shown in Figure 2 and Figure 3. As they show, in case of poolable data, the best performing estimator with the uniformly lowest MSE is the population mean or grand average estimator, as there is only one λ to estimate and the vector of Poisson variables is essentially a repetition of the same measurement. Also, of note is the performance of the zero estimator for small values of the true λ . For small values of the true λ , the zero estimator outperforms all estimators except for the population-average estimator. The reason the grand mean estimator performs best is that, in the case of poolable data, it is an unbiased estimator with variance, and hence, the MSE reduced proportionally to the number of measurements. We also expected a good performance from the zero estimator for small true λ values, as the shrinking target of the zero estimator is close to the true value, hence the estimator has a small bias and by definition a zero variance. The gain for the zero estimator over MLE is especially pronounced for the swab exposure, as in this case, the MLE by definition has a large variance. Notice that, for the swab exposure, the MLE starts to dominate the zero estimator for $\lambda_{\text{true}} > 500$ CFUs/m².

The second-best performing estimator for the swab exposure is EB-MOM, which has a significantly lower MSE than other estimators across almost the entire true λ range, except for small values of true λ s where it is dominated by the zero estimator. For the wipes, the pecking order of different estimators is similar, with the population mean estimator uniformly dominating all others. However, the region of the zero estimator's domination over MLE is much smaller, as the larger exposure of wipes reduces the MLE MSE. The EB-MOM and CZ estimators dominate the zero estimator for the whole range of true λ s while the performance of other shrinkage estimators is on par with the MLE. These results confirm that, for poolable data, the best strategy is to pool the data, as in this case, the pooled estimate has the smallest MSE.

The nonpoolable data set is presented in Figure 4 and Figure 5 for two types of exposures—swabs and wipes, respectively. The nonpoolable data sets were generated from a set of gamma distributions with increasing mean values and fixed variance equal to 10^5 for wipes and 10^6 for swabs. Figure 4 shows MSEs for the nine estimators for the wipe exposure and shows that, for nonpoolable data sets, the zero and population-average estimators are the worst performing due to the bias introduced by both estimators. On the other hand, the MLE and all shrinkage estimators perform well, with the Tsui estimator outperforming other estimators for small values of true λ s, as shown in the top insert in Figure 4. However, the second insert in Figure 4 shows that, for true λ values higher than 750 CFUs/m², the EB-MOM estimator starts to dominate all other estimators, including the Tsui estimator. Notice that the legend's color coding for the bottom insert in Figure 4 is different from the main figure and top insert. This result demonstrates that, for different true λ ranges, there might be different best estimators for nonpoolable data collected with wipes.

For the swab exposure, we used the same data-generating procedure except that the variance of the data-generating gamma distribution was set to 10^6 as the variance of the data collected with swabs is larger than for wipes. The results for the swab data simulations are shown in Figure 5. As can be seen from the figure, the Tsui estimator dominates all other estimators up to the true λ value being approximately 350 CFUs/m², after that, the EB-MOM takes over for the remaining range of λ . This result is similar to the result obtained for the wipe exposure except that the value where the two estimators flip is different.

To investigate the performance of shrinkage estimators on the data collected during the InSight mission, the data for the eight components shown in Table 1 were fitted with a gamma distribution that reflects the sample-to-sample variability of the data. The parameters of the fitted gamma distribution were $\alpha = 0.0447$ and $\beta = 2.5463 \cdot 10^{-4}$, with a λ average equal to 176 CFUs/m² and λ variance equal to $7 \cdot 10^5$ CFUs²/m⁴. The population variability distribution for the InSight data is shown in Figure 6. Having fit the data, the parameters of the population variability distribution generated true λ vectors of length 100, and those true λ s have been used to generate Poisson variables for a given exposure. This process has been repeated 100 times to quantify the uncertainty for each estimate. The performance of different shrinkage estimators for InSight data is shown in Figure 7 and Figure 8. As shown in those figures, the Tsui estimator dominates all other estimators for both types of exposure. This is in agreement with simulations shown on previous figures, as for relatively small true λ values, the Tsui estimator has the lowest MSE.

Figure 2: MSE as a Function of the True λ for Poolable Data with Swab Exposure Equal to 0.002 m².

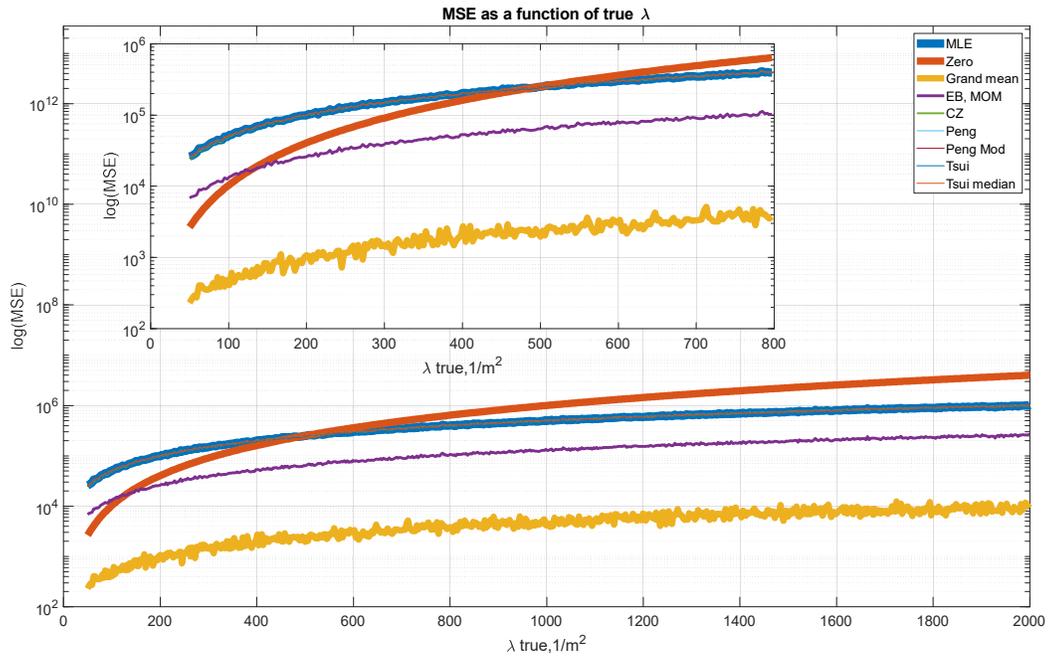


Figure 3: MSE as a Function of the True λ for Poolable Data with Wipe Exposure Equal to 0.25 m².

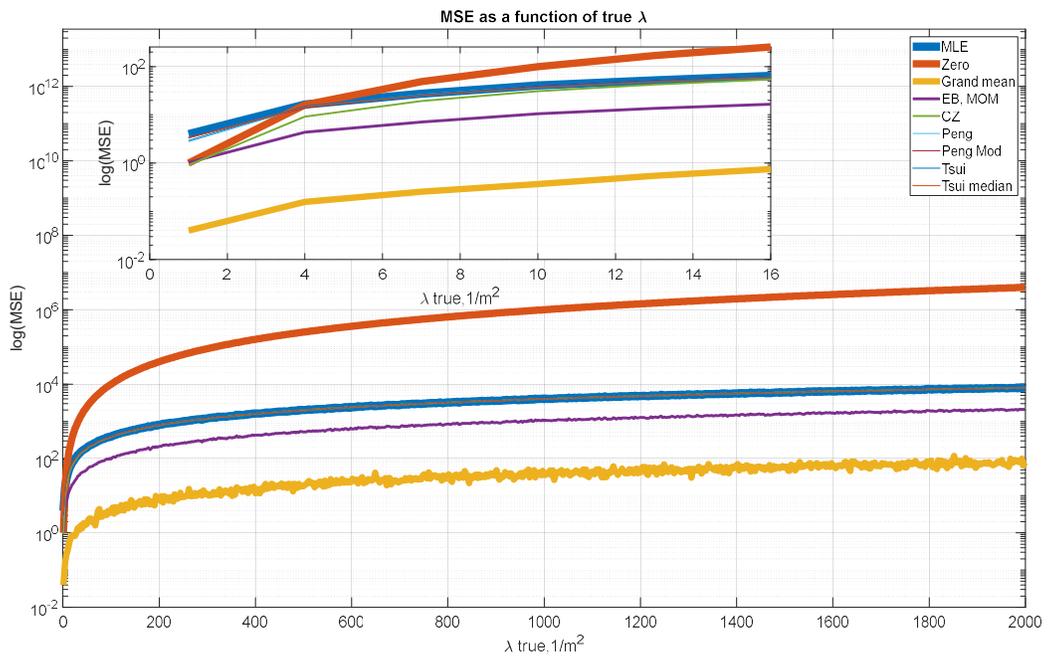


Figure 4: MSE as a Function of the True λ for Nonpoolable Data with Wipe Exposure Equal to 0.25 m². Bioburden Density λ s Were Generated from a Gamma Distribution with a Changing Distribution Mean Value While Keeping the Variance Constant at 10⁵.

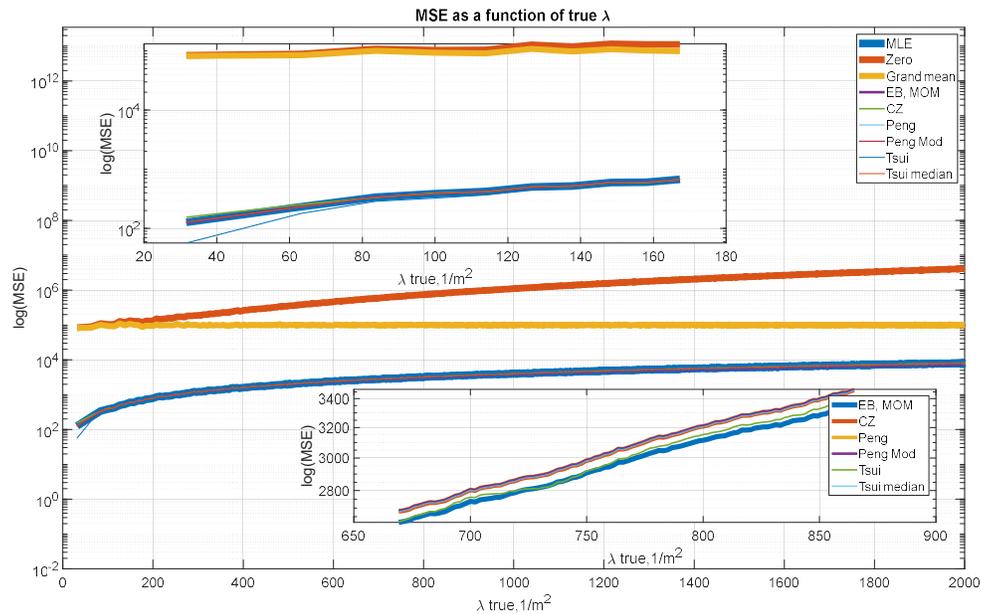


Figure 5: MSE as a Function of the True λ for Nonpoolable Data with Swab Exposure Equal to 0.002 m². Bioburden Density λ s Were Generated from a Gamma Distribution with a Changing Distribution Mean Value While Keeping the Variance at 10⁶.

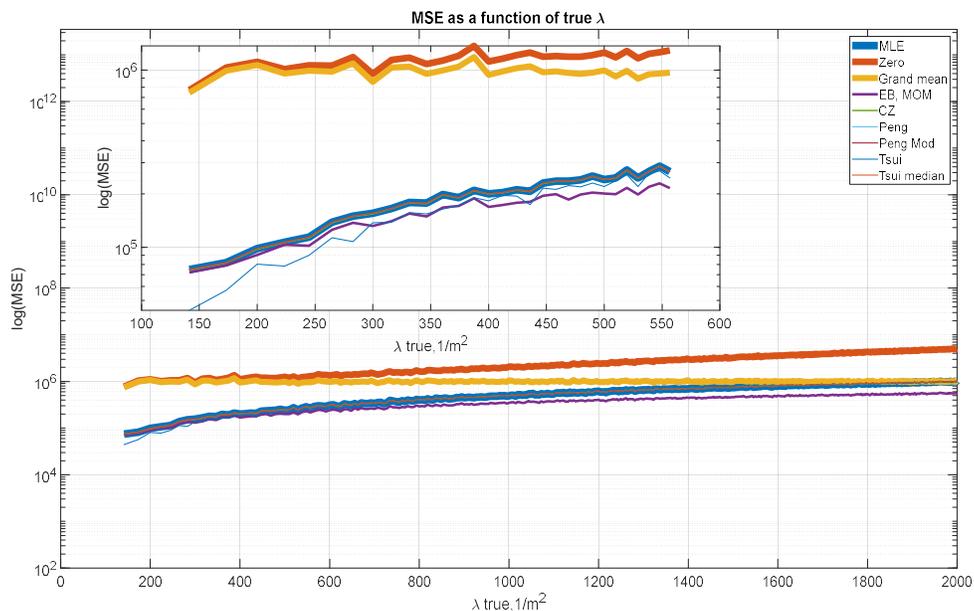


Figure 6: Sample-to-Sample Variability of InSight Data Approximated by a Gamma Distribution with Parameters $\alpha=0.0447$, $\beta=2.5463 \cdot 10^{-4}$, λ Average=175.5488 CFUs/m², and λ Variance=6.8943 $\cdot 10^5$ CFUs²/m⁴.

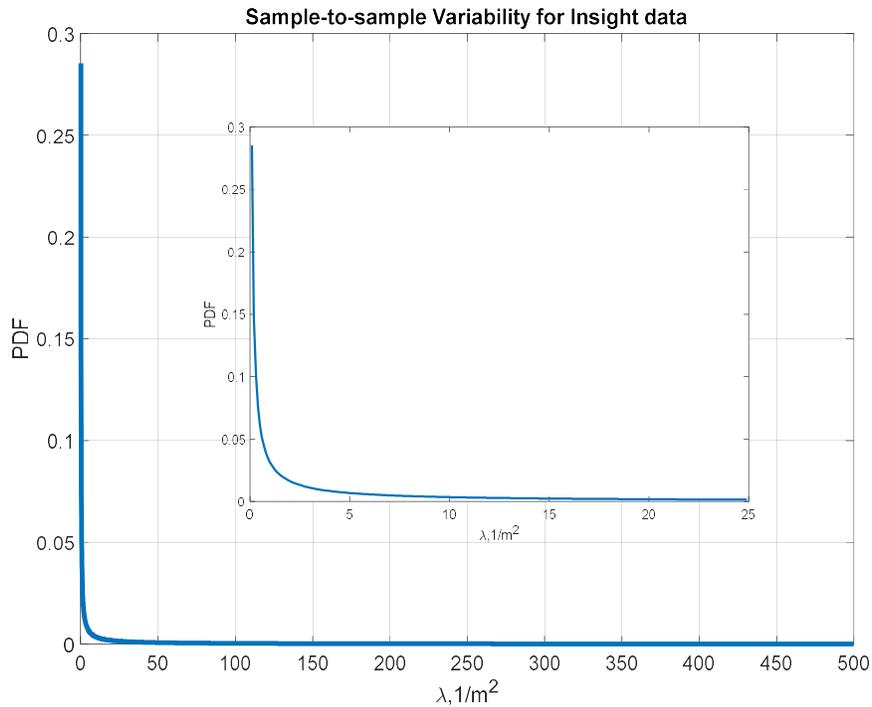


Figure 7: Performance of Different Shrinkage Estimators on a Simulated Gamma Distribution of InSight Data for Swab Exposure.

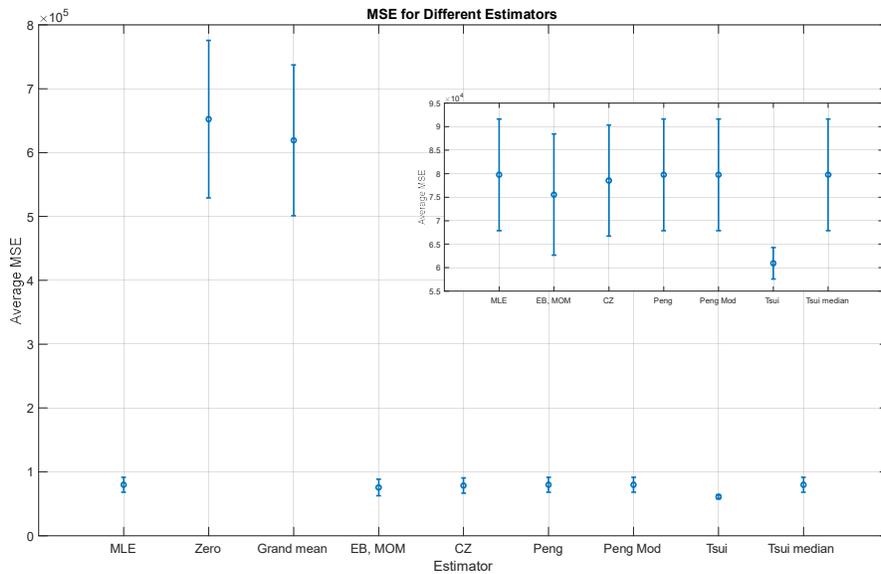
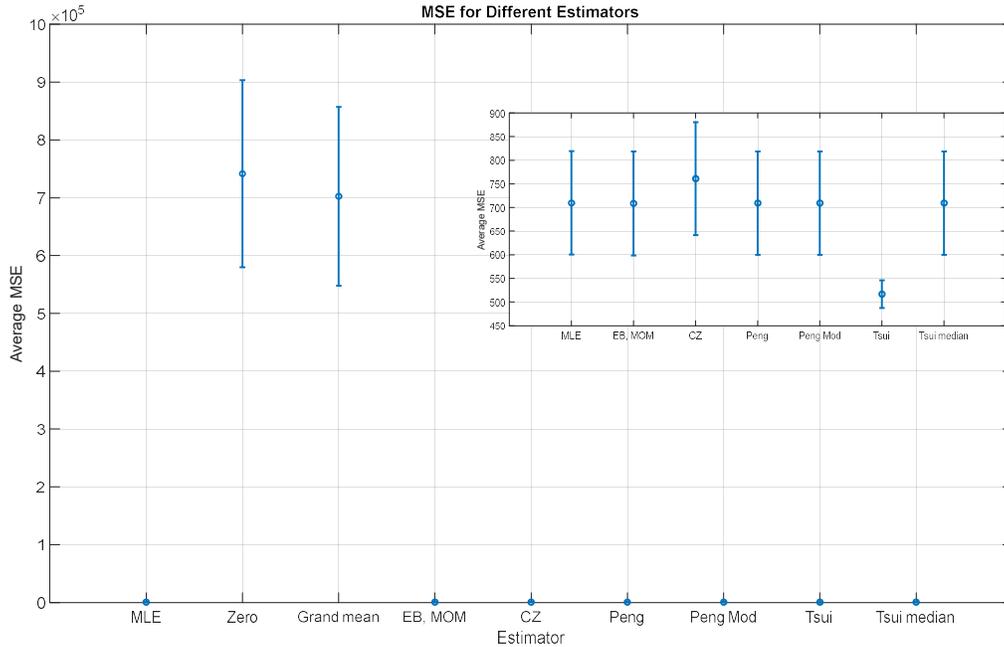


Figure 8: Performance of Different Shrinkage Estimators on a Simulated Gamma Distribution of InSight Data for Wipe Exposure.



4. CONCLUSION

We investigated the performance of nine different estimators on simulated data as well as the bioburden data collected during the InSight mission and used the data collected from eight different InSight components to generate test data to validate the estimators. The estimators’ performances were evaluated with respect to the mean squared loss function. The results show that, for poolable data sets, the best estimator is the population-average estimator, which uniformly dominates all other estimators in the relevant true λ range. For nonpoolable data sets, the estimators’ performances varied depending on the exposure. For the wipe exposure and true λ values less than 750 CFUs/m², the Tsui estimator dominates all other estimators; however, for larger true λ values, the EB-MOM estimator overtakes all other estimators by producing the lowest MSE. The same is true for the wipe exposure except for the fact that EB-MOM starts to dominate at true λ values around 350 CFUs/m². These results show that, for different true λ values, different estimators may be preferable. The future work will include comparison of shrinkage estimators with hierarchical Bayesian analysis.

For the data generated from the gamma population variability distribution obtained from the eight components of the InSight spacecraft, the estimator with the lowest MSE was the Tsui estimator, with all other estimators, except the population mean and zero estimators, having statistically similar MSEs independent of the exposure.

Acknowledgements

The research was partially carried out at the Jet Propulsion Laboratory, California Institute of Technology under a contract with the National Aeronautics and Space Administration (80NM0018D0004). Copyright © 2022 All rights reserved. We would like to thank the InSight mission team of Ryan Hendrickson, Gayane Kazarians, Lisa Guan, Sarah Cruz, and Pat Bevins for providing support in data generation. The authors would also like to thank Art Avila for their management support of this project and of this manuscript. This

research was supported by the National Aeronautics and Space Administration through the interagency collaboration agreement “Probabilistic Risk Analysis for Statistical Analysis of Bioburden Sampling Efficiencies” performed at Idaho National Laboratory. This manuscript has been authored by Battelle Energy Alliance, LLC under Contract No. DE-AC07-05ID14517 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for U.S. Government purposes

References

- [1] R. A. Beaudet. “*The statistical treatment implemented to obtain the planetary protection bioburdens for the Mars Science Laboratory mission,*” Adv. Space Research, volume 51, pp. 2261-2268, (2013).
- [2] J. N. Benardini, A. Seuylemezian, and A. Gribok, “*Predicting biological cleanliness: an empirical Bayes approach for spacecraft bioburden accounting,*” presented at the 2020 IEEE Aerospace Conference, 2020, pp. 1-12, <https://www.doi.org/10.1109/AERO47225.2020.9172725>.
- [3] A. Gribok, J. N. Benardini, and A. Seuylemezian. Bayesian Framework for Bioburden Density Calculations to Perform Planetary Protection Probabilistic Risk Assessment. Proceedings of the 30th European Safety and Reliability Conference and the 15th Probabilistic Safety Assessment and Management Conference. Edited by Piero Baraldi, Francesco Di Maio and Enrico Zio Copyright © 2020 by ESREL2020 PSAM 15 Organizers. Published by Research Publishing, Singapore ISBN: 981-973-0000-00-0 :: doi: 10.3850/981-973-0000-00-0 esrel2020p
- [4] A. Gribok, A. Seuylemezian, and J. N. Benardini. “*A Bayesian approach for estimating spacecraft microbial bioburden and managing the risk of biological contamination,*” presented at PSAM 2019, the Topical Conference Practical Use of Probabilistic Safety Assessment in Operations, December 2019, The Haymarket Hotel, Stockholm, Sweden.
- [5] H. F. Martz and R. A. Waller. “*Bayesian reliability analysis,*” reprinted with corrections. Krieger Publishing Co., 1991, Malabar, FL, USA.
- [6] C. L. Atwood, J. L. LaChance, H. F. Martz, D. J. Anderson, M. Englehardt, D. Whitehead, and T. Wheeler. “*Handbook of parameter estimation for probabilistic risk assessment,*” NUREG/CR-6823, SAND2003-3348P, U.S. Nuclear Regulatory Commission (2003).
- [7] B. Efron and C. Morris. “*Data analysis using Stein’s estimator and its generalizations,*” Journal of the American Statistical Association, volume 70, issue 350, pp. 311-319, (1975).
- [8] B. Efron and C. Morris. “*Stein’s estimation rule and its competitors--An empirical bayes approach,*” Journal of the American Statistical Association, volume 68, number 341, pp. 117-130, (1973).
- [9] G. Casella and R. L. Berger, “*Statistical inference.*” Duxbury, 2002, Belmont, CA.
- [10] L. Clevenson and J.V. Zidek. “*Simultaneous estimation of the means of independent Poisson laws,*” Journal of the American Statistical Association, volume 70, pp. 698-705, (1975).
- [11] J. L. Hodges and E. L. Lehmann. “*Some applications of the Cramer-Rao inequality,*” in Proceedings of the Second Berkeley Symposium, pp. 13-22, University of California Press, 1951, Berkeley.
- [12] J. Peng. “*Simultaneous estimation of the parameters of in-dependent Poisson distributions,*” Technical Report No. 78, Stanford University, Dept. of Statistics (1975).
- [13] K.-W. Tsui. “*Simultaneous estimation of several Poisson parameters under squared error loss,*” Technical Report No. 37, Stanford University, Dept. of Statistics (1978).