A Bayesian Approach for Estimating Spacecraft Microbial Bioburden and Managing the Risk of Biological Contamination

Andrei Gribok[a], Arman Seuylemezian[b], and J. Nick Benardini[b]
[a] Idaho National Laboratory, Idaho Falls, USA

[b] California Institute of Technology, Jet Propulsion Laboratory, Pasadena, USA

Abstract: Planetary protection (PP) is a discipline to minimize the inadvertent contamination of other planetary bodies by harmful microorganisms. The InSight mission was classified by NASA as a biologically sensitive mission to Mars. Thus, the flight hardware had to undergo microbial reduction and cleanliness verification/testing. NASA's approach since the 1970's provides a worst-case point estimation of the total microbial bioburden. This point estimation approach does not provide a probabilistic distribution of the bioburden, accounts for uncertainty arbitrarily and does not report credible intervals. To remediate these concerns, a Bayesian statistical approach was employed to estimate the microbial bioburden present on the InSight mission.

If no prior information about the microbial bioburden of the spacecraft is known, Bayesian analysis typically assumes a "non-informative" or weakly informative prior status, which can allow the data to essentially "speak for themselves." However, since Bayesian estimators are effectively shrinkage estimators, the effect of the initial non-informative status is most noticeable and potentially misleading when analyzing rare events. With very few non-zero data points, the shrinkage is particularly pronounced towards the mean value of the prior status as the likelihood provides no evidence that the bioburden density differs from zero.

This paper analyzes the performance of several non-informative priors—the Jeffreys non-informative prior, the Jeffreys constrained non-informative prior, and the Uniform prior and the Gamma (α,β) prior for different values of the parameters. The performance of different non-informative priors is evaluated in comparison to maximum likelihood estimation as well as through Bayesian model validation.

Keywords: NASA, Bayesian model, non-informative prior

1. INTRODUCTION

Probabilistic risk assessment (PRA) is a mature and influential technology that relies on two core methodologies—fault/event trees and statistical parameter estimation. The Contamination Probability Event Trees analysis utilizes Boolean logic to combine different paths to contamination events. Once such paths are exhaustively enumerated, the elementary probability rules are applied to aggregate the probabilities of different contamination scenarios into a contamination event. PRA performance depends critically on the accuracy of parameter estimates for individual components, as they are merged through event tree analysis to produce the overall probability of an inadvertent biological contamination event. One of the most important parameters affecting the probability of bio-contamination is the initial bioburden at launch, which accounts for the number of microorganisms present on the spacecraft. Historically, a frequentist approach has been used in planetary protection to estimate bioburden densities for individual components. However, in other PRA subject areas, the Bayesian approach has been used widely and successfully. We consider the two approaches complimentary rather than opposing each other because the Bayesian approach complements the frequentists' likelihood function with prior information in the form of distribution data against the parameter of interest. This paper compares and contrasts the two approaches used for these calculations and provides understanding of their performance using data collected during the InSight mission.

Planetary protection (PP) is an international discipline focused on the biological cleanliness of space exploration. The PP discipline was defined from the United Nations (UN) Outer Space Treaty of 1967, which provided the legal principles and framework for the exploration of outer space including that State Parties shall, "*conduct*

exploration of them so as to avoid harmful contamination and also adverse changes in the environment of Earth resulting from the introduction of extraterrestrial matter." In response to the UN Outer Space Treaty, the International Committee of Space Research (COSPAR) established a PP policy to avoid harmful biological contamination of celestial bodies from terrestrial sources, and harmful contamination of Earth from extraterrestrial sources referred to as forward and back PP, respectively. Each space-faring nation then developed their own set of requirements for spacecraft biological cleanliness based on target bodies and spacecraft type (e.g., flyby, orbiter, lander) and reports compliance to COSPAR on a biannual basis.

National Aeronautics and Space Administration (NASA) procedural requirement (NPR) 8020.12, "Planetary Protection Provisions for Robotic Extraterrestrial Missions," requires a stringent biological contamination prevention regime to planetary bodies of astrobiological interest for investigations of extant or evidence of extinct life, rather than those missions with no aims at life detection. For Mars missions, biological contamination concerns are the driver of hardware cleanliness requirements, which are satisfied by implementing biological reduction processes, recontamination prevention, and direct verification testing. As a measure of total microbiological cleanliness, NASA has selected the bacterial spore as a proxy for the total microbial bioburden for Mars-bound missions. The robust ability of the spores to resist harsh environmental conditions of interplanetary transfer and microbial reduction processes on the ground makes it the "worst case contaminant."

For a landed system not carrying instruments for investigations of extant life, the missions are required to not exceed 5×10^5 spores on launched spacecraft and limit the distribution of the total bioburden across the total sampled surface area of the spacecraft to an average bioburden density of ≤300 spores/m^2. To demonstrate compliance with these requirements, each project undergoes an extensive bioburden accounting effort, which tracks each individual hardware component throughout the assembly lifecycle, the associated biological reduction modalities implemented, and the bioburden density values acquired through direct testing of hardware components. NASA's technical specification for the verification of biological cleanliness on hardware surfaces (NASA Technical Handbook 6022) identifies the swab- and wipe-rinse methods as the predominate methodology to estimate the bioburden on spacecraft surfaces.

The hardware is sampled using a water-dampened swab or wipe throughout the integration and testing phase of the mission capturing a portion of the total surface area of each component. Once the sample is collected, it is then taken back to the lab where the sampling device is submersed in an extraction solvent and sonicated for 2 minutes ± 5 seconds to remove the biologicals from the sampling matrices. The extraction fluid is then heat-shocked at 80°C ± 2°C for 15 min. to select for heat-resistant organisms (e.g., spores). Aliquots of the extraction fluid are then placed into petri dishes ranging from 2–4 ml for swabs and 13–25 ml for wipes. A growth medium (tryptic soy agar) at 48–50°C is then added to the petri dishes and gently swirled to mix the extraction fluid and growth medium. After the mixture has cooled and settled, the petri dishes are aerobically incubated at 32°C and inspected for colony forming unit (CFU) at 24, 48, and 72 hours of incubation. The final raw CFU counts are recorded at 72 hours and are used for subsequent calculations. Associated hardware items are grouped together based on their assembly architecture and the raw CFU counts are used to calculate the bioburden density for a given hardware grouping. Given the previously established hardware hierarchy, the bioburden density of each hardware component contained within a group are then summed and extrapolated up to the entire spacecraft level. For a majority (>75%) of the groups, certain hardware components are deemed unsampleable and are instead accounted for using previously determined bioburden densities (i.e., the NASA specification values). Notably, for the purposes of this model development only hardware that was directly sampled was evaluated.

For Mars-destined spacecraft, the approach for performing raw CFU calculations to acquire bioburden densities have been negotiated between the NASA PP Office and each respective flight project on a mission-by-mission basis. The balance of these negotiations captures some combination of NASA's risk posture through conservatism from the policymakers with the existing science dataset and assumptions, sampling and processing efficiency

values, and the ability of a mission to extensively sample accountable surface areas while enabling projects to perform reasonable PP implementation strategies facilitating a successful launch.

The Mariner Mars 1971 and Viking missions used a sum-of-the-means mathematical approach that factored in raw CFU, sampling and processing efficiencies, surface area sampled, and total surface area. The Mars Pathfinder Mission (MPF) was the next mission to land on Mars in the late 1990s, which sent a table top-sized rover to Mars. Raw bioburden MPF counts were treated using a combination of Poisson and Gaussian statistics based on the sampling device, total CFU, and observed frequency of CFUs (e.g., many swabs over a surface with a low CFU probability). The statistical MPF treatment scenarios included Poisson statistics for swab samples with 0 CFU, Gaussian statistics for wipe samples with 0 CFU, and Gaussian statistics for a combination of wipe and swab samples with >1 CFU, all while factoring in the sum of the mean. Final bioburden density numbers were calculated using the sum of the 3-sigma value and either the true total CFU or an artificial value of 1 in the event of an observed total CFU of zero. In the early 2000s, golf cart sized rovers were developed and launched for the Mars Exploration Rover (MER) missions. MER employed an approach that was statistically identical to MPF. The next mission to Mars was the 2007–2011 Mars Science Laboratory (MSL) mission, which landed a spacecraft on the surface of the planet that was the size of a MINI Cooper automobile. Given the size of MSL, the previous statistics of MPF/MER would have treated all wipes in a Gaussian fashion, which would have been unreasonably conservative provided the mission used a majority of wipes to sample the inherently larger size of the hardware. Thus, MSL expanded the MER/MPF statistics to apply a Poisson statistical approach for wipe samples with 0 CFU and a Gaussian approach for cases where only one swab or wipe sample was taken and generated >1 CFU, referred to as the 3-sigma approach [2]. After the MSL launch, the NASA PP Office evaluated the approach for performing raw CFU calculations to acquire bioburden density and re-instated the Viking approach with the use of replacing the CFU value to an artificial count of 1 when a value of 0 CFU was observed for an entire group. This approach is referred to as the weighted averaging approach, which was used by the InSight mission from 2013–2018 and is being utilized on the current Mars 2020 mission as well.

Provided with the extensive dataset from multiple missions evaluating the different approaches of acquiring bioburden density utilized to date is key to help document and provide a technical mathematical and PP discipline rationale. The bioburden maximum approach has been good enough to meet bioburden requirements despite the introduction of unnecessary error and loosely defined mathematical substitutions of CFU. These past missions have been afforded the flexibility to negotiate statistical implementation approaches in regard to a mission's biosampling verification approach, reported bioburden maximum values, and employed a NASA policy direction to change the CFU from 0 to 1 while still meeting their bioburden requirements. However, as future missions involving the exploration of Outer Planets, sample return, and carrying humans to Mars become more complex, a maximum bioburden density that is overly conservative may not allow for requirements compliance. Therefore, for missions with probabilistic risk requirements, such as the upcoming Europa Clipper mission, bioburden distribution is a key parameter (N_0) that defines the initial starting bioburden contamination for the system [3]. Given a more realistic understanding of bioburden on the risk assessment of the spacecraft can help to not only satisfy mission requirements with greater confidence, but also drive microbial reduction trades and ensure overkill treatments are not unnecessarily applied to hardware resulting in possible negative impacts and potentially additional cost and schedule to projects. The development of this approach will allow PP to: (1) account for inherent errors; (2) understand the distribution rather than a point estimate of spores on the spacecraft; and (3) monitor individual subsystems with associated confidence levels. By taking past mission use cases into account and implementing a baseline Bayesian modeling approach as a means to estimate the bioburden density from the raw CFU counts, it will help to establish both a technically robust mathematical and biological approach using actual spacecraft test data to evaluate a suite of real-life spacecraft life cycle scenarios.

The data generated by a PP verification campaign encompasses 3–5 years, >10 distinct geographic locations with 2–3 critical cleanrooms, and >3,000 direct verification tests resulting in ~40,000 petri dishes. The InSight mission collected 2,031 swabs and 1,266 wipes on spacecraft hardware surfaces over nearly 4 years. Given the clean spacecraft, 93% of the swabs and 63% of the wipes had a 0 CFU count at 72 hours resulting in ~85% of the

39,379 petri dishes yielding 0 CFU. Several use cases from the InSight mission were utilized herein for model development and validation, but this dataset is representative of the MER and MSL datasets as well with >95% of the swabs and >60% of the wipes from direct verification testing yielding 0 CFU at 72 hours. For the InSight dataset, a 1 was added to 127 groups when a 0 CFU result was observed yielding 2,557 artificial spores, a 1.7% increase to the reported final total mission bioburden.

NASA specification values are assumed for hardware components that cannot be sampled and are based on the manufacturing and test cleanroom environments and associated process controls. Specification values are utilized in the planning stages of the mission for hardware allocation values when no prior engineering judgment exists. These specifications can then be verified by direct sampling of a given component at the vendor site or upon receiving inspection. For example, the surface microbial density for hardware in a highly controlled International Standards Organization (ISO) Class 5 cleanroom is 50 spores/m^2 as compared to a standard ISO Class 8 cleanroom at 1×10^4 spores/m^2. To incorporate this pre-determined conservative estimate into the non-informative priors evaluated in this work, a constrained non-informative (CNI) was developed that allowed the incorporation of a pre-determined expected bioburden density of 300 spores/m^2 derived from the average bioburden density requirement for the InSight mission. This allowed for a policy-based engineering value to be represented in the form of a non-informative prior distribution, which would ultimately influence the bioburden density estimations. This non-informative prior can thus be utilized by space missions in the future as an absolute worst-case value or by NASA headquarters (HQ) for verification activities. Future iterations of this non-informative prior information might consider incorporating other expected values of the bioburden density based on various cleanroom environmental conditions to better reflect an unknown environment or a well-known environment compared to the default conservative approach used in this work.

2. DATA COLLECTION AND PREPROCESSING

Several different use cases were utilized from the InSight mission to compare the application of various non-informative prior distributions on the final bioburden density distributions obtained. In addition, raw data from the same components were used to generate the currently approved and utilized weighted average approach, as well as the previously utilized 3-sigma approach for comparison. The data for each component were collected using either swabs or wipes. For each component, a number of samples were collected on one given date or on several different dates. A swab data collection covered the area of 0.0025 m^2 with a single swab, while a wipe covered area varied typically between 0.1 and 1.0 m^2 depending on the geometric complexity and size of the sampled component. Each individual swab or wipe was considered a sample. Having been processed in the microbiology laboratory, the samples were deposited in petri dishes and covered with tryptic soy agar. For swabs, only 80% of the total sample solution was deposited in the dishes, thus producing a pour fraction of 0.8 that was taken into account by reducing the surface area sampled (exposure). For wipes, the pour fraction was 0.25. The number of samples varied for each component could have been sampled either with swabs, wipes, or both depending on the size and complexity of the component. Only true observed counts were utilized for the Bayesian approach; no NASA policy directives were employed to change the counts from 0 to 1 afterwards. For the purpose of Bayesian analysis, the raw data for each component were represented by pairs (x_i,e_i), i=1,2...N, where x_i is the number of 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. Data for each component were pooled to produce a total count and a total exposure as $X=\sum_1^N x_i$ and $E=\sum_1^N e_i$. The total count and total exposure have been used in Poisson likelihood for Bayesian inference. “Exposure” and “effective sampled area” are used in this paper interchangeably. Eight different InSight components were selected for analysis and represent a range of different total CFU counts, effective areas sampled, and total areas of the components. Table 1 summarizes this data.

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

Component	CFU Count	Area Sampled, m^2	Exposure: Area Sampled*Pour Fraction, m^2	Total Surface of the Component, m^2	% Sampled=Area Sampled/Total Area
9	0	0.6031	0.2167	0.7580	79.5650
73	0	2.4200	0.6160	2.7400	88.3212
300	1	2.6600	0.6705	5.0000	53.2000
169	1	0.2400	0.1920	0.5850	41.0260
283	5	4.5710	1.1427	12.0000	38.0920
243	5	0.2800	0.1140	0.2980	93.9600
38	12	3.1050	0.8065	10.0000	31.0500
261	52	0.0600	0.0480	0.3120	19.2310

3. BAYESIAN INFERENCE WITH NON-INFORMATIVE PRIORS

The specification of prior distribution is one of the most important methodological, as well as practical, problems in Bayesian inference. While the major appealing property of Bayesian inference is its ability to include prior information into the inference model, sometimes it is required to avoid reliance on old information as it may dominate the newly collected data, especially if the new data are sparse. Another reason to do away with informative priors is the lack of reliable information about the parameter of interest. Finally, an informative prior may be considered "subjective" by peers; as such, using a non-informative prior may help to alleviate this concern.

To address all three issues for the bioburden calculations mentioned above, we performed Bayesian inference using a Gamma-Poisson compound distribution model, shown in Eq. (1), with four different non-informative prior distributions. The aleatory model assumes that CFUs are distributed on the surface of the spacecraft according to a Poisson distribution with its single parameter—λ specifying bioburden density. The bioburden density is measured in CFU/m^2 and is the main parameter of interest in this study.

$$P(\lambda/X) = \frac{\underbrace{\frac{(\lambda\cdot E)^X}{X!}e^{-\lambda\cdot E}}_{Likelihood}.\underbrace{\frac{\beta^{\alpha}.\lambda^{\alpha-1}.e^{-\lambda\cdot\beta}}{\Gamma(\alpha)}}_{Prior}}{\int_0^{\infty}\underbrace{\frac{(\lambda\cdot E)^X}{X!}e^{-\lambda\cdot E}}_{Likelihood}.\underbrace{\frac{\beta^{\alpha}.\lambda^{\alpha-1}.e^{-\lambda\cdot\beta}}{\Gamma(\alpha)}}_{Prior}d\lambda} \quad (1)$$

where λ is bioburden density, E is exposure, X is current total CFU count data, α and β is shape and rate parameters of Gamma distribution, and Γ(α) is Gamma function. The empirical evidence for CFUs being distributed according to a Poisson distribution was first reported in [8]. Since then, the Poisson distributional model is widely used to represent rare events occurring either over time periods or surfaces and volumes.

The reasons for applying different non-informative priors to the same datasets are twofold: (1) to study the sensitivity of posterior and predictive inference to non-informative priors; and (2) to attempt to select the "best" non-informative prior for the problem in hand, which can be used as a "default" non-informative prior. The four non-informative priors that were used in this paper are: (1) Jeffreys non-informative prior (Jeffreys) [1,5]; (2) constrained noninformative prior (CNI) [1]; (3) Maximum Entropy prior (MaxEnt) [4]; and (4) Uniform prior (Uniform) [9]. Several reasons motivated the selection of these specific non-informative priors. First, they all can be parameterized as a Gamma function, which makes them conjugate, thus significantly simplifying computations and allowing interpretation of posterior parameters in terms of data. Second, Jeffreys and Uniform priors are the most widely used non-informative priors in Bayesian inference; however, for Poisson likelihood, they are

improper, thus complicating Bayesian model selection. They also shrink the parameter of interest to values that can be unrealistic. In addition, while Jeffreys is invariant to re-parameterization, Uniform is not. Third, CNI and MaxEnt priors can explicitly include information about the moments of prior distribution, thus providing more leverage over posterior and predictive inference. The choice of non-informative prior is especially important in cases of rare events when many data samples contain zeros and the posterior inference relies exclusively on prior assumptions.

The Jeffreys non-informative prior uses the Fisher information matrix to place maximally non-informative prior data on the parameters, exploiting the fact that the Fisher information matrix is widely considered as an indicator of the accuracy of a parameter estimate. For the Gamma-Poisson model, the Jeffreys prior can be parametrized as Gamma(0.5,0), which is an improper distribution; however, it can produce the proper posterior, as long as the exposure is not zero.

Uniform prior distribution can be parameterized as Gamma(1,0) and is also improper while producing a proper posterior. In this paper, it is defined in the domain as $0<\lambda<2000$.

One of the disadvantages of using improper prior distributions is the lack of well-defined moments for those distributions, so it is difficult to foresee how the distribution will affect the data. In this sense, while being defendable and simple, improper distributions could be quite unrealistic since the posterior mean is pulled towards a mean value of a prior distribution. To address this shortcoming, two proper non-informative priors have been applied, which both can explicitly specify the expected mean value of the bioburden.

The first prior was the MaxEnt prior, which is an extension of Laplace's principle of indifference stating that under the lack of evidence, all possible outcomes should be considered equiprobable. For a discrete random variable, Shannon [9] introduced the notion of informational entropy, which can be used as a measure of "randomness." In the case of discrete random variables, the uniform distribution happens to have the largest entropy; hence, it is considered the most random. The uniform distribution does assign equal probabilities to all outcomes, thus satisfying the principle of indifference. For continuous random variables with a probability density function of $p(\theta)$, the definition can be extended and is called differential entropy in Eq (2).

$$H = -\int p(\theta)\log\left(p(\theta)\right) d\theta \tag{2}$$

However, in contrast to its discrete counterpart, differential entropy can be infinite, negative or positive, and non-invariant under the change of variables. Most importantly, there is no single maximum entropy distribution for continuous variables as differential entropy depends on constraints placed on distributional moments and the support of a variable. Since maximum entropy can only be calculated for continuous distributions under constraints, placing constraints on a random variable's expected value and defining it in the positive half-line makes exponential distribution a maximum entropy distribution. Exponential distribution can be parameterized as Gamma(1,1/μ) where μ is a pre-defined expected value.

The lack of invariance under reparameterization motivated development of a CNI prior [1] which allows specifying the expected value of the parameter in addition to retaining desirable properties of the maximum entropy distribution. In contrast to differential entropy, the CNI uses the definition of entropy as suggested in [4], which is the negative of Kullback–Leibler divergence with a reference distribution $\pi(\theta)$ being the Jeffreys prior in Eq. (3):

$$H = -\int p(\theta)\log\left(\frac{p(\theta)}{\pi(\theta)}\right) d\theta \tag{3}$$

In addition to the reparameterization property, the CNI also has a larger variance than the MaxEnt, thus diminishing the influence of the prior and giving more weight to the data. The variance of prior distribution plays

a prominent role in Bayesian inference as it defines how strongly the posterior mean is "pulled" towards the prior mean. The CNI can be parametrized as Gamma(1/2,1/(2·μ)), where μ is the pre-defined expected value. Table 2 provides a summary of the four non-informative priors used in this paper. The posterior distribution mean has been used as posterior summary statistics. The 90% credible intervals were used to quantify uncertainty in posterior inference. For this paper, μ=300 CFU/m^2 was used to reflect the allocation based on NASA requirements, as nearly all bioburden controlled hardware should be at or below this value post-microbial reduction in order to receive a hardware certification for flight.

Table 2: Parameters of Gamma Distribution for Different Non-Informative Priors.

Prior distribution	α- shape	β - rate	Mean	Variance
Jeffreys	0.5	0	undefined	undefined
CNI	0.5	1/(2·μ)	μ	2·μ^2
MaxEnt	1	1/μ	μ	μ^2
Uniform	1	0	undefined	undefined

Along with the bioburden density λ, NASA requires an estimate of the total number of CFU for each spacecraft at launch. This required the evaluation of the total CFU numbers for each component before summing them up. To perform this estimation, posterior predictive Bayesian inference has been applied using posterior predictive distributions for the Gamma-Poisson model in Eq. (4):

$$\mathrm{NB}\left(\tilde{x}; X, \alpha_{post}, \underbrace{\frac{\beta_{post}}{\beta_{post}+\tilde{A}}}_{p}\right) = \int_0^{\infty} \underbrace{\frac{(\lambda\cdot\tilde{A})^{\tilde{x}}}{\tilde{A}!} e^{-\lambda\cdot\tilde{A}}}_{Likelihood} \cdot \underbrace{\frac{{\beta_{post}}^{\alpha_{post}}\cdot\lambda^{\alpha_{post}-1}\cdot e^{-\lambda\cdot\beta_{post}}}{\Gamma(\alpha_{post})}}_{Posterior} d\lambda \tag{4}$$

where NB is a negative binomial distribution, α_{post} and β_{post} are the parameters of Gamma posterior distribution, $\tilde{A}$ is the total area of the component, and $\tilde{x}$ are the predicted CFU values.

The posterior predictive inference is summarized as the mean value of a posterior predictive distribution and 90% credible intervals. Bayesian model selection has been performed using the Bayes factor (BF) approach, which relies on the comparison of posterior odds for different models and the selection of a model, which is most supported by the observed data. If there are two competing models represented by two different prior distributions parameterized as Gamma (α_1, β_1) and Gamma (α_2, β_2) for each model, then the marginal data likelihood is calculated as (Eq. 5):

$$P(X/\alpha_i, \beta_i) = \int_0^{\infty} \underbrace{\frac{(\lambda\cdot E)^X}{X!} e^{-\lambda\cdot E}}_{Likelihood} \cdot \underbrace{\frac{{\beta_i}^{\alpha_i}\cdot\lambda^{\alpha_i-1}\cdot e^{-\lambda\cdot\beta_i}}{\Gamma(\alpha_i)}}_{Prior} d\lambda, i = 1,2 \tag{5}$$

The BF can then be calculated as (Eq. 6):

$$BF = \frac{P(X/\alpha_1,\beta_1)}{P(X/\alpha_2,\beta_2)}, \tag{6}$$

which is the ratio of total probability of observed data to occur under two different models.

Credible intervals were calculated as the inverse of the Gamma distribution function [7].

4. RESULTS AND DISCUSSION

Tables 3–10 summarize the posterior and predictive inference for the components listed in Table 1.

Table 3: Summary of Posterior and Predictive Inference for Component 9.

Prior Distribution	Posterior Mean Bioburden Density – λ, CFU/m^2	5th Percentile of Posterior Distribution	95th Percentile of Posterior Distribution	Predictive Mean, CFU	5th Percentile of Predictive Distribution	95th Percentile of Predictive Distribution
Jeffreys	2.3064	0.0090	8.8600	1.7482	0	7
CNI	2.2889	0.0090	8.7928	1.7350	0	7
MaxEnt	4.5432	0.2330	13.6102	3.4437	0	11
Uniform	4.6128	0.2366	13.8189	3.4965	0	11

Table 4: Summary of Posterior and Predictive Inference for Component 73.

Prior Distribution	Posterior Mean Bioburden Density – λ, CFU/m^2	5th Percentile of Posterior Distribution	95th Percentile of Posterior Distribution	Predictive Mean, CFU	5th Percentile of Predictive Distribution	95th Percentile of Predictive Distribution
Jeffreys	0.8116	0.0031	3.1180	2.2239	0	9
CNI	0.8094	0.0031	3.1096	2.2180	0	9
MaxEnt	1.6146	0.0828	4.8370	4.4241	0	14
Uniform	1.6233	0.0832	4.8631	4.4479	0	14

Table 5: Summary of Posterior and Predictive Inference for Component 300.

Prior Distribution	Posterior Mean Bioburden Density – λ, CFU/m^2	5th Percentile of Posterior Distribution	95th Percentile of Posterior Distribution	Predictive Mean, CFU	5th Percentile of Predictive Distribution	95th Percentile of Predictive Distribution
Jeffreys	2.2371	0.2623	5.8274	11.1855	1	30
CNI	2.2315	0.2617	5.8130	11.1579	1	30
MaxEnt	2.9680	0.5273	7.0401	14.8404	2	37
Uniform	2.98280	0.5299	7.0750	14.9140	2	37

Table 6: Summary of Posterior and Predictive Inference for Component 169.

Prior Distribution	Posterior Mean Bioburden Density – λ, CFU/m^2	5th Percentile of Posterior Distribution	95th Percentile of Posterior Distribution	Predictive Mean, CFU	5th Percentile of Predictive Distribution	95th Percentile of Predictive Distribution
Jeffreys	7.8120	0.9162	20.3497	4.5700	0	13
CNI	7.7452	0.9083	20.1757	4.5309	0	13
MaxEnt	10.2389	1.8192	24.2859	5.9897	0	15
Uniform	10.4161	1.8507	24.7063	6.0934	0	16

Table 7: Summary of Posterior and Predictive Inference for Component 283.

Prior Distribution	Posterior Mean Bioburden Density – λ, CFU/m^2	5th Percentile of Posterior Distribution	95th Percentile of Posterior Distribution	Predictive Mean, CFU	5th Percentile of Predictive Distribution	95th Percentile of Predictive Distribution
Jeffreys	4.8129	2.0016	8.6086	57.7549	22	105
CNI	4.8059	1.9987	8.5961	57.6713	22	105
MaxEnt	5.2352	2.2799	9.1730	62.8226	26	112
Uniform	5.2504	2.2865	9.1996	63.0053	26	113

Table 8: Summary of Posterior and Predictive Inference for Component 243.

Prior Distribution	Posterior Mean Bioburden Density – λ, CFU/m^2	5th Percentile of Posterior Distribution	95th Percentile of Posterior Distribution	Predictive Mean, CFU	5th Percentile of Predictive Distribution	95th Percentile of Predictive Distribution
Jeffreys	48.2413	20.0632	86.2868	14.3759	5	28
CNI	47.5504	19.7758	85.0510	14.1700	4	27
MaxEnt	51.1363	22.2700	89.5997	15.2386	5	29
Uniform	52.6269	22.9191	92.2115	15.6828	5	30

Table 9: Summary of Posterior and Predictive Inference for Component 38.

Prior Distribution	Posterior Mean Bioburden Density – λ, CFU/m^2	5th Percentile of Posterior Distribution	95th Percentile of Posterior Distribution	Predictive Mean, CFU	5th Percentile of Predictive Distribution	95th Percentile of Predictive Distribution
Jeffreys	15.4988	9.0584	23.3428	154.9887	88	237
CNI	15.4671	9.0398	23.2949	154.6710	88	236
MaxEnt	16.0526	9.4952	24.0081	160.5268	92	243
Uniform	16.1188	9.5343	24.1070	161.1883	93	244

Table 10: Summary of Posterior and Predictive Inference for Component 261.

Prior Distribution	Posterior Mean Bioburden Density – λ, CFU/m^2	5th Percentile of Posterior Distribution	95th Percentile of Posterior Distribution	Predictive Mean, CFU	5th Percentile of Predictive Distribution	95th Percentile of Predictive Distribution
Jeffreys	1093.5221	857.6728	1353.0301	341.1789	262	428
CNI	1057.0469	829.0645	1307.8988	329.7986	253	414
MaxEnt	1032.4675	810.7817	1276.2804	322.1298	248	404
Uniform	1103.9366	866.9054	1364.6267	344.4282	265	432

As can be seen from Tables 3–6, the posterior mean values are segregated into two groups—one consists of Jeffreys and CNI priors, while the other consists of MaxEnt and Uniform. This is primarily due to the fact the α-shape parameters are the same for the Jeffreys and CNI, and the MaxEnt and Uniform, while the β-rate parameter is mostly affected by the total exposure and not the prior parameter value. Since the total count is very small or zero for components 9, 73, 300, and 169, the shape parameter is hardly impacted at all, thus producing two groups of priors with different posterior mean values. As evident from Tables 7–10, as the total number of counts increases, the difference in posterior means between the two groups decreases. The same observation is true for all other values reported in Tables 3–10 as the number of total counts increases, the posterior and predictive inference becomes similar for all priors used in this paper. This reflects a fundamental property of Bayesian inference-prevalence of the data over the prior as the amount of collected data increases. This observation also suggests that selection of a "default" best prior is more important for scarce datasets then for more complete datasets. The 90% credible intervals get wider as the number of counts increase. This is probably because the likelihood function gets wider since for the datasets with the large counts, there is more uncertainty about the value of the bioburden density. Predictive credible intervals also correlate with the percentage of sampled area as evidenced from Tables 7 and 8 representing components 283 and 243. Component 243 was sampled in its entirety; hence, its credible predictive intervals are much narrower than that for Component 283, only 38% of which was sampled. The mean values of predictive distribution also correlate with the total CFU count found on the component.

The pair-wise comparison of priors for Components 73 and 261 is presented in Tables 11 and 12. It should be noted that the BFs do not select the "right" model, rather they indicate how much the data favors one model over the other. For the pair-wise comparison, we assume the priors in the rows of Tables 11 and 12 are the null hypothesis, while the priors in the columns represent the alternative hypothesis. Diagonals in both tables have values of 1 since the hypothesis is compared to itself. From Table 11, it can be seen that CNI has BFs higher than 1 with respect to all other competing priors. According to the Jeffreys' scale [6], a BF higher than 5 offers substantial evidence in favor of a null hypothesis, one higher than 10 gives strong evidence in favor of a null hypothesis, and a BF >150 indicates decisive evidence in favor of null. According to this scale, CNI is strongly favored over the Jeffreys prior for component 73, substantially favored over MaxEnt, and decisively favored over Uniform prior distribution for the same component. For component 261, the domination of CNI is significantly diminished, as seen in Table 12. CNI is weakly favored over Jeffreys and MaxEnt, but substantially favored over Uniform. This demonstrates diminishing importance of prior selection for components with substantial CFU counts as the data starts to dominate the prior.

Table 11: Pair-wise BFs for Component 73.

Prior	Jeffreys	CNI	MaxEnt	Uniform
Jeffreys	1	0.0770	0.5491	181.9690
CNI	12.8842	1	7.0758	2344.5426
MaxEnt	1.8208	0.1413	1	331.3423
Uniform	0.0054	0.0004	0.0030	1

Table 12: Pair-wise BFs for Component 261.

Prior	Jeffreys	CNI	MaxEnt	Uniform
Jeffreys	1	0.4598	0.5633	5.4081
CNI	2.1748	1	1.2251	11.7617
MaxEnt	1.7751	0.8162	1	9.6003
Uniform	0.1849	0.0850	0.1041	1

Since BF model selection seems to favor CNI over other priors, in Table 13 we present a comparison of estimates obtained using this prior with NASA's current and legacy approaches. The legacy 3-σ approach is a purely frequentist approach based on the assumption of Gaussian statistics for CFU distribution and estimation of the mean value of this distribution and its standard deviation, σ [2]. Having estimated the mean value, the 3σ is added to it for conservatism. As can be seen from Table 13, this approach systematically produces bioburden values higher than the Bayesian approach with a CNI prior. On the other hand, the current weighted average approach employs an ad-hoc Bayesian procedure equivalent to using the Uniform prior. In contrast to the Uniform prior approach used in this paper, the weighted average uses one more parameter (sample device efficiency) in calculation of the effective area. Sampling device efficiency is not a directly observable value, it is an estimate, so it had been chosen not to use it in the Bayesian approach described in this paper. While the 3-σ approach envisages the calculation of confidence intervals and weighted average does not, neither technique currently reports them. In comparison to Bayesian analysis, both techniques generally produce higher values of bioburden density, thus ultimately reporting a higher bioburden density at launch and/or resulting in a higher probability of inadvertent biological contamination when used in PRA models.

Table 13: Comparison of Bayesian Posterior Inference Using CNI Prior with NASA Weighted Average and 3-σ Approaches for the Eight Components.

	Proposed Bayesian Approach			NASA Legacy	NASA Current
Component	Posterior Mean Bioburden Density – λ, CFU/m^2	5th Percentile of Posterior Distribution	95th Percentile of Posterior Distribution	3 sigma Bioburden Density – λ, CFU/m^2	Weighted Average Bioburden Density – λ, CFU/m^2
9	2.2889	0.0090	8.7928	13.84	27.99
73	0.8095	0.0032	3.1097	4.87	17.36
300	2.2315	0.2617	5.8130	5.96	9.54
169	7.7452	0.9083	20.1757	20.83	33.70
283	4.8059	1.9987	8.5961	5.17	11.11
243	47.5504	19.7758	85.0510	130.14	186.70
38	15.4671	9.0398	23.2949	52.06	9.66
261	1057.0469	829.0645	1307.8988	2349.53	658.47

It is also instructive to analyze the influence of different priors on the data as presented in Figure 1. It can be seen that the Jeffreys and Uniform priors are the least informative in the sense that the posteriors produced using these priors are the closest to the likelihood function. Meanwhile, the CNI and MaxEnt priors are the two prior samples producing posteriors most distinct from the likelihood. In this sense, these two priors could be called mildly informative. Note that the MaxEnt mode is the furthest from the likelihood function, reflecting the fact that the MaxEnt has the smallest variance among the four; hence, it pulls the posteriors more strongly toward the prior mean value. In comparison to historically generated bioburden values, the mathematical approach employed for the InSight mission (weighted average) and the MSL mission (3-sigma) was utilized to calculate the bioburden density of the components found in Table 13. Briefly, the weighted average approach implemented on the InSight mission and currently the approved methodology for conservatively performing bioburden calculations assumed an artificial CFU count of 1, for components where the true count was 0. It then considered the same efficiency and pour fraction values utilized to perform the Bayesian posterior inferences and generated bioburden density values based on the sum of the raw counts from a given sample set for each sampling date. These individual

densities representing data from samples taken on a single date were then combined into a weighted average that was reported as the bioburden density for a given component. The MSL approach utilized a 3-sigma approach described in [2] to generate bioburden densities. This approach was replicated using the raw data from each respective component provided in Table 13. Generally, the calculated bioburden density using either the 3-sigma or weighted average approach tended to be a more conservative estimate than that generated using the Bayesian methodology. However, neither the weighted average nor the 3-sigma approach provided a distribution of bioburden densities and instead reported the maximum upper bound. In addition, these techniques arbitrarily applied conservatism to increase confidence in the estimation and overestimate the true bioburden. Although these techniques were appropriately implemented during past missions, they do not provide the necessary robustness and mathematical rigor for the increasing complexity of upcoming missions. The Bayesian mathematical approach not only provides a robust methodology to calculate the distribution of bioburden and associated confidence intervals present on spacecraft surfaces, it also allows the incorporation of engineering judgement, as well as appropriate conservatism in the estimations, thereby facilitating compliance of more stringent and sensitive bioburden requirements.

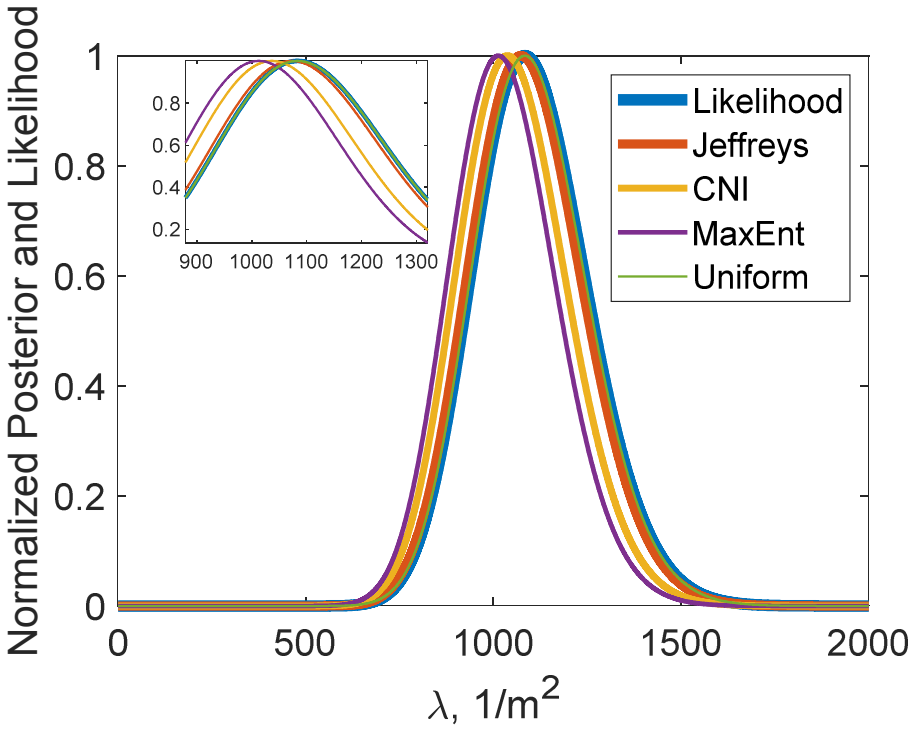

Figure 1. Normalized Posterior Densities and Likelihood Function for Component 261.

4. CONCLUSIONS

In the absence of substantial prior information, a common approach is to apply a non-informative prior and then proceed with the utilization of a posterior inference. However, the selection of a "default" non-informative or weakly informative prior is a problem specifically as it depends on the likelihood model, data and required inference uncertainty. In this paper, we analyzed, validated and contrasted four different non-informative priors, which can be used to establish a spacecraft bioburden value that can be used to verify at launch bioburden requirements for a Mars-destined spacecraft or used to provide an initial bioburden input into the PRA model to assess inadvertent spacecraft biological contamination. The priors have been applied to the bioburden data collected for the InSight mission. Components with different counts, sampled areas, and total areas were selected to validate different priors. In addition, bioburden densities generated using Bayesian analysis have been compared with the current (weighted average) and legacy (3-sigma) bioburden calculation approaches used in past

NASA missions. It was found that non-informative priors play a significantly more prominent role in posterior inference for components with low CFU count data. As the number of counts increase, the prior's influence decreases. However, the uncertainty of posterior inference increases as the number of counts also increases, thus reflecting uncertainty in the data with more counts. The same observation is true for predictive inference. Bayesian model selection was performed using pair-wise comparison of BFs for different priors. Based on the presented data, the CNI prior seems to be the most appropriate prior to use as a "default" prior for the bioburden data. As expected, the Jeffreys and Uniform priors proved to be the least informative; however, BF favored CNI and MaxEnt priors over them. Comparison with current and legacy methods used by NASA to calculate bioburden revealed that both methods tend to overestimate bioburden density, and hence, likely overestimating the final reported bioburden and the probability of bio-contamination in the risk assessment models. Overestimating the risk may lead to additional cleaning sessions, thus increasing the cost and hardware reliability of the mission and may discredit efforts to comply with the increasingly stringent cleanliness requirements for the missions of tomorrow. Using non-informative priors is just one approach to perform Bayesian inference; the others include the empirical Bayes approach and hierarchical Bayes, both of which will constitute future work.

Acknowledgements

The research described in this publication was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA. 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 and Melissa Jones for their management support of this project and critical read of this manuscript.

References

1. Atwood, C. L., 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).
2. Beaudet, R. A. "*The statistical treatment implemented to obtain the planetary protection bioburdens for the Mars Science Laboratory mission*," Adv. Space Research, vol. 51, pp. 2261–2268 (2013).
3. DiNicola, M., K. McCoy, C. Everline, K. Reinholtz, and E. Post. "*A mathematical model for assessing the probability of contaminating Europa*," 2018 IEEE Aerospace Conference, Big Sky, MT, USA, pp. 1-20 (2018).
4. Jaynes, E. T. "*Prior probabilities*," IEEE Trans. Systems Sci. Cybernetics, vol. SSC4, pp. 227–241 (1968).
5. Jeffreys, H. "*An invariant form for the prior probability in estimation problems*," Proceedings of the Royal Society of London Series A, Mathematical and Physical Sciences, vol. 186 (1007), pp. 453–461 (1946).
6. Jeffreys, H. *The Theory of Probability* (3rd ed.). Oxford, England, p. 432 (1998).
7. Martz, H. F., and R. A. Waller. "Bayesian Reliability Analysis," reprinted with corrections. Krieger Publishing Co., Malabar, FL, USA (1991).
8. Neyman, J. "*Lectures and conferences on mathematical statistics and probability*," Graduate School, U.S. Dept. of Agriculture, Washington DC. Available at: http://hdl.handle.net/2027/mdp.39015007297982 (last accessed September 16, 2019) (1952).
9. Shannon, C. E., and W. Weaver. "*The Mathematical Theory of Communication*," University of Illinois Press (1949).