Modeling of Recovery Efficiency of Sampling Devices used in Planetary Protection Bioburden Estimation

Michael DiNicola[a], Arman Seuylemezian[b], Christine Moissl-Eichinger[c], Amy Baker[d], and Jason Johns[e]

[a] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, USA, Michael.Dinicola@jpl.nasa.gov

[b] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, USA, Arman.Seuylemezian@jpl.nasa.gov

[c] Medical University of Graz, Diagnostic and Research Institute of Hygiene, Microbiology, and Environmental Medicine, Graz, Austria, christine.moissl-eichinger@medunigraz.at

[d] SETI Institute, Mountain View, USA, amy.l.baker@nasa.gov

[e] National Space and Aeronautics Administration, Kennedy Space Center, Merritt Island, USA, jason.w.johns@nasa.gov

---

Abstract: Microbial contamination has been of concern to the planetary protection discipline since the Viking missions in the 1970s. Spacecraft missions destined for other planetary bodies must abide by a set of requirements put forth by National Aeronautics and Space Administration (NASA) based on recommendations from the Committee on Space Research (COSPAR). Compliance with these biological cleanliness requirements are demonstrated by direct sampling of spacecraft hardware and associated surfaces to enumerate the number of microorganisms present on the surface. The discipline has employed a variety of tools to perform direct sampling, including four types of swabs (Cotton Puritan, Cotton Copan, Polyester, and Nylon Flocked) and two types of wipes (TX3211 and TX3224), which are typically used to sample surfaces no larger than 25 cm^2 and 1 m^2, respectively. The recovery efficiency of these devices is a critical parameter used to generate spacecraft level cleanliness estimates.

In this study, we investigate how recovery efficiency differs by inoculum amount and species. This is analysed across different sampling devices using a set of microbial organisms applied to stainless steel surfaces. Two different assaying techniques were employed: the NASA standard assay and the European Space Agency (ESA) standard assay; as well as two different techniques for plating: Milliflex filtration and direct plating. Data were analysed by first developing a probabilistic model of the experimental process, from coupon inoculation through recovery of colony forming units. The model quantifies the probability that an individual microorganism is recovered, a key metric for predicting bioburden. A cost function is developed to identify those assay methods that provide optimal bioburden estimation capability. Results suggest the nylon flocked swab and the TX3211 wipe yielded the highest recovery efficiency and optimal bioburden estimation capability. Results from this study will be integrated into a larger statistical framework assessing planetary protection bioburden requirements.

---

1 INTRODUCTION

In compliance with the international treaty put forth by Committee on Space Research (COSPAR) as well as agency level requirements imposed by National Aeronautics and Space Administration (NASA) and the European Space Agency (ESA). Both the biotechnology and planetary protection group at the Jet Propulsion Laboratory (JPL) and respective groups at ESA have been monitoring the microbial bioburden of spacecraft and associated surfaces to minimize the inadvertent forward contamination of other planetary bodies and preserve the scientific integrity of exploratory missions. The primary devices used for acquiring samples of spacecraft hardware are cotton swabs and polyester wipes. These samples are then subjected to laboratory processing to enumerate the bioburden density and total microbial bioburden present on a given

surface [1]. The NASA standard assay which outlines the laboratory processing procedures primarily select for hardy heat-tolerant microorganisms which are used as a proxy for the total microbial load present on the surface [2].

Direct sampling takes place throughout the entire assembly lifecycle of the mission through to launch [3]. As such a variety of surface materials are sampled and a diverse set of microorganisms are typically recovered [4]. However, due to operational, logistical and budgetary constraints, planetary protection engineers collect samples from a subset of the entire spacecraft surface. Surface counts are then extrapolated to estimate the total microbial bioburden and the bioburden density of the entire spacecraft in order to demonstrate compliance to launch requirements.

In order to accurately estimate the microbial bioburden present on the spacecraft surface, it is imperative to quantify the efficiency with which the sampling device recovers microbial organisms. Previous studies aimed at quantifying the sampling device efficiency focused on a single device [5]. However, no studies to date have examined the range of sampling devices currently in use in the planetary protection discipline. In this comprehensive study, we examined all swabs and wipes currently approved for planetary protection use. In addition, we developed a model to capture uncertainties present throughout the entire experimental process including seeding of the stainless-steel coupons, sampling of the surfaces and wet laboratory processing. In addition, to understand the sensitivity of each sampling device to recovering different bacterial species, a range of species commonly recovered from spacecraft surfaces were used, primarily belonging to the most commonly isolated *Bacillus* genera.

Due to mitigation protocols that minimize microbial contaminants, typical spacecraft surface samples are extremely clean and non-zero colony forming units (CFU) are sparse. As such these experiments, focused on quantification of the device recovery efficiency with lower inoculum amounts to mimic real-world conditions. Although a wide range of surface materials are sampled throughout the life cycle of each unique mission, stainless steel has been established in several previous studies as a representative spacecraft material , hence it was the surface material of choice in this study.

Previous studies have demonstrated a statistical framework for performing bioburden accounting [6, 7], however they have not directly accounted for the recovery efficiency of the sampling devices used [8]. The results communicated in this study will be integrated with the statistical framework currently under development [8] and will ultimately be used for performing bioburden accounting. Directly accounting for the recovery efficiency of these assay methods will provide to provide a more accurate estimate of the total microbial bioburden and bioburden density originating from planetary protection sampling efforts. Finally, we use the term “assay methods” throughout to refer to the process of acquiring a sample from a surface using a given sampling device, storing the sample and delivering it to the lab, and extracting, plating and observing CFU. The assay method for swabs is pictured in the CFU recovery portion of Figure 1.

2 MATERIALS AND METHODS

2.1 Laboratory Facilities

Experiments included in this study were conducted by two research groups in two distinct laboratories. All wipe experimental data was generated through NASA at Kennedy Space Center facilities. Swab data was generated through ESA at the Medical University Graz Center for Microbiome Research. Additionally, experiments using the Puritan cotton swab (the primary swab employed by NASA), were also conducted through NASA at the Kennedy Space Center facilities.

2.2 Sampling Devices

Four different swab types were used in this study, including the Copan (Murrietta, CA) polyester (PE) ATP-free (170C) plain swab, Copan nylon flocked swab (552C), Copan cotton swab (150CA) and the Puritan (Guilford, ME) cotton swab (806 WC). Both the Copan PE swab and the Copan cotton swab arrived gamma sterilized from the manufacturer. The Puritan cotton swab was sterilized by autoclaving, and the Copan nylon flocked swab was sterilized by EO15.

Two different wipes were used in this study including the TexWipe (Kernersville, NC) TX3211 and TX3224 polyester wipes. Wipes were prepared for sampling by folding and rolling them into a 50 mL conical tube, 10-20 mL of sterile water was added to saturate the wipe. The tube is then lightly capped and sterilized by autoclaving. The wipes are then allowed to cool to room temperature prior to sampling.

2.3 Preparation of Spores

A total of six different organisms were used in this study including *Bacillus (B.) atrophaeus* DSM 675, *B. safensis* DSM 19292, *B. megaterium* DSM 32, *B. megaterium* 2c1, *B. thuringensis* DSM 2046 and *B. thuringensis* E24. Spores from all six organisms were prepared following protocols highlighted in [9]. Spore stocks were stored in 50% ethanol.

2.4 Seeding of Spores

For swab experiments, stainless steel coupons from (Wilms Metallmarkt Lochnleche, Koln, Ehrenfeld, Germany) and 1.5 mm x 50.5 mm x 50.5 mm in size were used. Stainless steel coupons were sterilized for 3 hours at 160 °C. Various concentrations of spores were spotted onto coupons in droplets containing 4 µL to minimize spread. Coupons were then left to dry under laminar flow for 24 hours. For consistent precision amongst recovery experiments conducted by both NASA and ESA, the number of replicates was varied based on the initial inoculum amount and expected recovery, shown in Table 1. Note that the experiments performed by ESA with the Copan cotton swab and by NASA with the Puritan swab did not include the case where the targeted inoculation level was 3 CFU due to limits of detection in the experiment. All swab experiments testing the sensitivity of recovery efficiency to inoculum amount used the species *B. atrophaeus*. In both NASA and ESA conducted swab experiments, sensitivity of recovery efficiency to species was addressed at the 100 CFU concentration.

For wipe experiments, stainless steel metal coupons were manufactured by the KSC Prototype Development Lab they were 16 in x 16 in x 0.04 inch in size were used. Coupons were autoclaved to ensure sterility. Various concentrations of spores were then spotted onto coupons in 4 µL droplets. Coupons were then left to dry overnight under laminar flow. As with swab experiments, the number of replicates was varied based on the initial inoculum amount as shown in Table 1, and the sensitivity of wipe recovery efficiency to inoculum amount was tested using the species *B. atrophaeus*. To assess the sensitivity of recovery efficiency of wipes to species, experiments were performed at a 400 CFU concentration.

2.5 Recovery of Spores using Swabs

Two different approaches were taken in processing the swabs, one followed the NASA standard protocol as outlined in NPR 8715.24, the other followed the ESA standard protocol outlines in ECSS-Q-ST-70-55C.

For the ESA standard protocol, after vortexing and sonication, the sample solution containing the swab head was typically plated onto two separate R2A plates using a spread plate technique and 1.0 mL of solution was transferred to each plate. In some instances, four plates with 0.5 mL solution in each plate were used to accommodate higher inoculation levels or when higher recovery efficiency was achieved. Plates were subsequently incubated at 32 °C and colony counts were performed at 24, 48 and 72 hours.

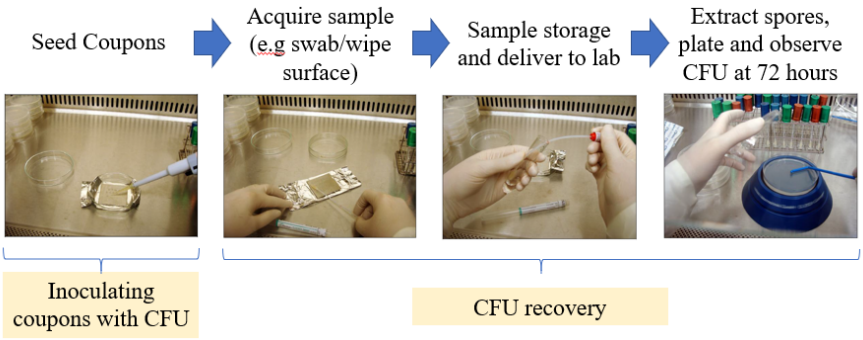

Figure 1. The seeding and recovery process used in the swab recovery efficiency experiments of this study, described in Section 0. This is the data generating process that is modelled in Section 0.

For the NASA standard protocol, after vortexing and sonication the sample solution was plated onto two sterile petri dishes, 4 mL of solution was used in each plate using a pour plate technique. Plates were then left out to dry and subsequently incubated at 32 °C. Colony were counted at 24, 48 and 72 hours.

In addition to experimental conditions, several controls were used. Negative controls were performed by sampling sterile stainless-steel coupons, positive controls were performed by direct plating of the spore solutions for each inoculation level assessed.

2.6 Recovery of Spores using Wipes

The wipes were processed using only the NASA standard protocol as outlined in NPR 8715.24 (NASA, 2021). After suspending the wipe in planetary protection rinse solution containing 85 mg/liter potassium dihydrogen phosphate, 200 mg/liter Tween 80; pH 7.2 and subjecting it to sonication and vortexing. the solution was then filtered through a 0.45 micron PVDF membrane filter using the Millipore milliflex membrane filtration system (Merck KGaAm Darmstadt, Germany) and plated onto Millipore filter cassettes (#RMHVMFX24) containing TSA. At first, approximately half of the wipe solution was poured through the filter. The remaining volume was then sonicated and vortexed for 5-10 seconds before pouring the solution through a second filter. The resulting plates were then incubated at 32 °C, colony counts were then performed at 24 and 48 hours.

Table 1. Summary of sampling devices, processing techniques and testing facilities used in this study. For each of these cases, the target inoculation levels tested and respective number of replicates performed for each inoculation level are shown to the right.

Sampling Device	Processing Technique	Testing Facility	Inoculation Level	# of Replicates
Copan PE Swab	ESA Standard	ESA	3, 5, 10, 15, 100	50, 40, 30, 20, 10
Nylon Flocked Swab	ESA Standard	ESA	3, 5, 10, 15, 100	50, 40, 30, 20, 10
Puritan Cotton Swab	NASA Standard	ESA	3, 5, 10, 15, 100	50, 40, 30, 20, 12
Copan Cotton Swab	NASA Standard	ESA	5, 10, 15, 100	40, 30, 20, 10
Nylon Flocked Swab	NASA Standard	ESA	3, 5, 10, 15, 100	20, 10, 10, 10, 10
Puritan Cotton Swab	NASA Standard	NASA	5, 10, 15, 100	40, 30, 20, 10
TexWipe TX3211	NASA Standard/Milliflex Filter	NASA	16, 50, 160, 400	16, 8, 8, 8
TexWipe TX3224	NASA Standard/Milliflex Filter	NASA	16, 50, 160, 400	16, 8, 8, 8

3 A MATHEMATICAL MODEL OF RECOVERY EFFICIENCY

The mathematical model we develop to study recovery efficiency of various sampling devices and protocols considers the end-to-end process that generates the CFU data observed from each experiment, as discussed in Section 0 and illustrated in Figure 1. The experimental process is partitioned into two sub-models: an initial seeding model, formulated in Section 3.1; and a recovery model, formulated in Section 3.2.

Importantly, the recovery model is dependent on the outcome of the seeding model. Not capturing this dependency tends to ignore the uncertainty in the number of CFU seeded onto the coupon, using a single number equal to the average of control experiments for the number of CFU seeded onto the coupon. This, as we will see in Section 4, has ramifications on how certain we are of a given method's recovery efficiency. Finally, a cost function is developed in Section 3.3 to quantitatively compare methods as to their effectiveness in bioburden estimation applications. Stan software [10] in RStudio [11] were used to develop models and figures in software.

3.1 Seeding Model

The experimental process described in Section 2.4 begins with the seeding of a coupon with a targeted number of CFU, which we refer to as the target inoculation level. As shown in [12], this process satisfies necessary independence and uniformity properties to be modeled by a Poisson distribution. In this study, we perform a rigorous statistical analysis of how λ varies with the target inoculation level using observations from control experiments performed.[1]

Let I be the set of target inoculation levels of interest and J_i be the total number of control experiments performed at the target inoculation level $i \in I$. The probability that we observe $\mathrm{n_{i,j}}$ CFU in the j^{th} control experiment for the target inoculation level i given the parameter value λ is $p\left(n_{i,j} \mid \lambda(i)\right) = e^{-\lambda(i)} \frac{\lambda(i)^{n_{i,j}}}{n_{i,j}!}$, where λ follows a power law with respect to the target inoculation level; $\ln\big(\lambda(\mathrm{i})\big) = \delta_0 + \delta_1 \ln(i)$. Since the experimental design takes precautions to isolate coupons when seeding (e.g. biosafety hoods) and implements processes to avoid cross-contamination between experiments (e.g. usage of sterile pipettes between coupon inoculations) all observations of the number of CFU seeded onto each coupon are taken to be independent of one another. Therefore, the probability of observing all control experiment results $\mathbf{n} = \left(n_{i,j}\right)_{j=1,i\in I}^{J_i}$ given the parameter values (δ_0, δ_1), is

$$p(\,\mathbf{n} \mid \delta_0, \delta_1\,) = \prod_{i\in I} \prod_{j=1}^{J_i} p\left(n_{i,j} \mid \lambda(i)\right). \tag{1}$$

The parameters (δ_0, δ_1) are given the joint prior distribution $p(\delta_0, \delta_1) = p(\delta_0) \times p(\delta_1)$, where the prior distributions of δ_0 and δ_1 are independent normal distributions with means $\mu_{\delta,0}$ and $\mu_{\delta,1}$, and standard deviations $\sigma_{\delta,0}$ and $\sigma_{\delta,1}$, respectively. Independence is assumed in the prior for simplicity and computational purposes, as there is enough data to uncover dependencies during fitting of the model. Since, prior to performing control experiments, the number of CFU initially placed on a surface is expected to be centered around the target inoculation level, we set $\mu_{\delta,0} = 0$, $\mu_{\delta,1} = 1$. We also set $\sigma_{\delta,0} = 5$ and $\sigma_{\delta,1} = 1$. Sensitivity analysis and prior predictive checks demonstrated that the model results were not significantly affected by a broad range of reasonable values for these parameters.

Using Bayes' theorem to calculate the posterior distribution $p(\delta_0, \delta_1|\boldsymbol{n})$ from Equation (1), the posterior predictive distribution of the number of CFU placed on a new coupon not in the control experiments, such as a coupon used to test recovery efficiency, when the target inoculation level is some positive integer $\hat{\imath}$, is given by

$$p(\,\tilde{n} \mid \hat{\imath}, \boldsymbol{n}\,) = \int_{\delta_1=-\infty}^{\infty} \int_{\delta_0=-\infty}^{\infty} p\big(\,\tilde{n} \mid \lambda(\hat{\imath})\,\big)\; p(\delta_0, \delta_1 |\, \boldsymbol{n})\; d\delta_0\; d\delta_1. \tag{2}$$

Note that it may be the case that $\hat{\imath} \notin I$ and so Equation (2) can be used to predict the number of CFU present on a coupon when other inoculation levels are targeted besides those used in the control experiments.

[1] Control experiments performed for these recovery efficiency experiments inoculated growth medium directly.

3.2 Recovery Model

The experimental process described in Section 2.4 begins with the seeding of a coupon with a targeted number of CFU, which we refer to as the target inoculation level. Suppose that a coupon has been inoculated with n CFU of a specified species as discussed in Section 3.1 after targeting inoculation level $i \in I$, and that the probability of recovering an individual CFU using a specific assay method is θ. As with the seeding process, the design of the recovery experiments takes several measures to ensure uniformity of the sampled surface, sampling device and assay protocol and to avoid outside contamination: coupons are of the same material type, size and from the same manufacturer; reagents are tested for sterility according to manufacturer recommendations, and sampling devices are applied to surfaces and CFU are extracted from these devices, plated in growth medium, cultivated and counted following prescribed protocols (NASA or ESA standard assays). These measures help ensure that all CFU have the same probability of recovery, θ for a given assay method. Moreover, sonication and vortexing when seeding and during extraction minimize such phenomenon as microorganism clumping, which allows us to model CFUs as being recovered independently of one another. With these assumptions in place, the probability that r CFU are recovered from a total of n CFU on a coupon is given by

$$p(\, r \mid n, \theta\,) = \binom{n}{r}\theta^{r}(1-\theta)^{n-r}, \tag{3}$$

which is recognized as the binomial distribution. This distribution has the property that the expected value of the number recovered $E[R] = \sum_{r=0}^{n} r\; p(\, r \mid n, \theta\,)$ equals $n \times \theta$, or equivalently, $\theta = \frac{E[R]}{n} = E\left[\frac{R}{n}\right]$, which is the mean recovery efficiency. Experimental observation of the number recovered allow us to estimate $E[R]$ by a sample mean $\bar{r}$ which, by the Law of Large Numbers, will converge to $E[R]$ as the sample size increases, implying that $\theta \approx \frac{\bar{r}}{n}$. Hence, when n is known, the probability that an individual CFU is recovered can be approximated by the mean recovery efficiency.

In order to consider potential over-dispersion in this study's recovery data, we will allow θ to be realizations from a beta distribution parameterized by a mean value μ and dispersion parameter φ:

$$p(\, \theta \mid \mu(i), \varphi\,) = \frac{\Gamma(\varphi)}{\Gamma(\mu(i)\varphi)\Gamma\left((1-\mu(i))\varphi\right)}\theta^{\mu(i)\varphi-1}(1-\theta)^{(1-\mu(i))\varphi-1}, \tag{4}$$

where Γ is the gamma function. In this study, we would also like to assess whether or not a trend exists between the mean recovery efficiency and the target inoculation level. To do this, we further let μ be a function of γ_0 and γ_1 using the logit transformation, $\mu(i) = \frac{1}{1+e^{-(\gamma_0+i\gamma_1)}}$, which has found broad applicability in bio-assay research [13]. Hence, given that a coupon is originally seeded with n microorganisms after targeting an inoculation level of i, the probability that r CFU are recovered using a specified assay method is $\int_{\theta=0}^{1}\binom{n}{r}\,\theta^{r}(1-\theta)^{n-r}\,p(\, \theta \mid \mu(i), \varphi\,)\; d\theta$, which solves to become

$$p(\, r \mid n, \mu(i), \varphi\,) = \binom{n}{r}\frac{\Gamma(\varphi)}{\Gamma(\mu(i)\varphi)\Gamma\left((1-\mu(i))\varphi\right)}\frac{\Gamma(r+\mu(i)\varphi)\Gamma\left(n-r+(1-\mu(i))\varphi\right)}{\Gamma(n+\varphi)}\;, \tag{5}$$

which is recognized as the beta-binomial distribution.

The number seeded onto a coupon does not depend on the recovery method or protocol as is made explicit by its lack of dependence on the parameters $\mu(i)$ and φ. Moreover, Equation (2) gives us a model of the number of CFU seeded onto a coupon used for the recovery experiments when the inoculation level is i. Hence, we have that $p(\, n \mid \mu(i), \varphi\,) = p(n)$ and let $p(n) = p(\, \tilde{n} \mid \hat{\imath}, \mathbf{n}\,)$. Therefore, the probability that r CFU are recovered using a specified assay method given parameter values $\mu(i)$ and φ is $p(\, r \mid \mu(i), \varphi\,) = \sum_{n=r}^{\infty} p(n)\; p(\, r \mid n, \mu(i), \varphi\,)$.

Turning to the actual experiments performed using a given assay method, for each inoculation level targeted $i \in I$ and each of M_i independent experiments performed at that inoculation level, there were a total of $r_{m,i}$ CFU observed in recovery. The probability of recovering $\mathbf{r} = \left(r_{i,m}\right)_{m=1,i\in I}^{M_i}$ across all independently performed experiments is

$$p(\mathbf{r} \mid \gamma_0, \gamma_1, \varphi) = \prod_{i\in I} \prod_{m=1}^{M_i} p\left(r_{i,m} \mid \mu(i), \varphi\right) \tag{6}$$

Recall that $\mu(i)$ is a function of the parameters γ_0 and γ_1. The parameters $(\gamma_0, \gamma_1, \varphi)$ are given the joint prior distribution $p(\gamma_0, \gamma_1, \varphi) = p(\gamma_0) \times p(\gamma_1) \times p(\varphi)$. Here, γ_0 and γ_1 are given normal distributions and φ is given an exponential distribution with rate parameter κ (note that dispersion is strictly positive). For the same reasons discussed in Section 3.1, independence is assumed in the joint prior distribution even though dependence is known to exist among regression coefficients. When designing the experiments, the mean recovery efficiency was assumed to be 50% to first order, with a broad degree of uncertainty. Given this information prior to doing the experiments, we set the means for γ parameters $\mu_{\gamma,0} = \mu_{\gamma,1} = 0$ so that the mean of $\mu(i)$ is $\frac{1}{2}$ and set $\sigma_{\gamma,0} = 5$ and $\sigma_{\gamma,1} = 1$ to allow for considerable variability as to the value of $\mu(i)$ and the trends it allows with the target inoculation level. The value of κ was set to $\frac{1}{1000}$ to allow the data to more strongly inform the dispersion parameter, φ. Sensitivity analysis demonstrated that the model results were not significantly affected by a broad range of reasonable values for these parameters.

Using Bayes' theorem to calculate the posterior distribution $p(\gamma_0, \gamma_1, \varphi \mid \mathbf{r})$ from Equation (6), the posterior distribution of θ when the inoculation level is $\hat{\imath}$ is calculated by integrating over all possible values of $\mu(\hat{\imath})$ and φ:

$$p(\theta \mid \hat{\imath}, \boldsymbol{r}) = \int_{\gamma_1=-\infty}^{\infty} \int_{\gamma_0=-\infty}^{\infty} \int_{\varphi=0}^{\infty} p(\theta \mid \mu(\hat{\imath}), \varphi)\ p(\gamma_0, \gamma_1, \varphi \mid \mathbf{r})\ d\gamma_0\ d\gamma_1\ d\varphi. \tag{7}$$

Note that we have used the fact that γ_0, γ_1 and φ do not depend on $\hat{\imath}$, and that, once $\mu(\hat{\imath})$ and φ are known, no further information is provided by the data $\mathbf{r}$ when determining the distribution of θ. Finally, the 95% credibility interval from the marginal posterior distribution of γ_1 was used to test for significance of a trend between recovery efficiency and inoculation level. If this interval contained 0, then it was concluded that there is insufficient evidence to include a trend parameter in the model at this time. Otherwise, it was concluded that the dependency between recovery efficiency and inoculation level should be further investigated.

3.3 Bioburden Estimation

Probabilistically estimating bioburden utilizes much of the theory developed in sections 3.1 and 3.2. A sampling device, such as a swab or wipe, and assay method such as the NASA standard assay, are applied to sample a surface. The sampling device is then contained in a sterile container and transported to a lab for extraction, plating and culture. After a specified amount of time, CFU are counted and recorded. Key differences between this process and the one described in the previous sections are

1. In real applications, there can be many different species of microorganisms present on surfaces, with differing recovery characteristics.
2. The number of microorganisms on the surface prior to sampling is not experimentally controlled, but is dependent on the cleanroom's air flow properties, human activity in the cleanroom and other phenomenon that are not well understood at this time.

To consider (1), we introduce the notion of a microorganism ``recovery type". Two individual microorganisms are of the same recovery type if they share the same probability of being recovered. Let π_k, for $k = 1, \ldots, K$, represent the probability that an individual microorganism is of recovery type k, and let θ_k be the probability that an individual microorganism is recovered given it is of type k. For example, two species of microorganism might be considered of the same recovery type if the difference between their probabilities of being recovered is statistically insignificant as judged by the 95% credibility interval. Since current assay protocols do not identify microorganism species upon culture, the probability that r CFU are recovered from Equation (3) conditioned on knowing the number of microorganisms present on the surface and the probabilities of an individual microorganism being of a given recovery type becomes

$$p(r \mid n, \pi_1, \ldots, \pi_K, \theta_1, \ldots, \theta_K) = \binom{n}{r}\left(\sum_{k=1}^{K} \pi_k \theta_k\right)^r \left(1 - \sum_{k=1}^{K} \pi_k \theta_k\right)^{n-r} \quad (8)$$

where Equation (8) equals zero when $r > n$, and $\pi_k \in [0,1]$ for all k such that $\sum_{k=1}^{K} \pi_k = 1$.[2] In practice, we typically reserve the last recovery group K for the set of "novel" microorganisms that have unknown recovery efficiency probability relative to the sampling device or protocol being applied, and $\pi_K = 1 - \sum_{k=1}^{K-1} \pi_k$. The recovery probability for an individual microorganism of a given type k is $p(\theta_k \mid n, \mathbf{r}_k)$ as calculated by Equation (7), where $\mathbf{r}_k$ is the vector of experimental observations of the recovery efficiency of recovery type k microorganisms relative to a given assay method. When $k = K$, it will be assumed that θ_k follows the distribution in Equation (4) with $\mu = \frac{1}{2}$ and $\varphi = 1$, also referred to as a Jeffreys' Prior distribution.

To consider (2), this study assumes that cleanroom fallout of microorganisms onto surfaces follows a Poisson distribution with mean parameter, λ. We assume ignorance of the correct Poisson model, and so we let λ follow a distribution proportional to $\lambda^{-\frac{1}{2}}$ although a more physics-based model will better inform this probability. Under these assumptions, the probability that there are n microorganisms on the surface is $p_0(n) \propto \frac{\Gamma\left(n+\frac{1}{2}\right)}{\Gamma(n+1)}$ prior to the observation of recovery data.[3]

Continuing to assume that individual microorganism recoveries are independent, the probability that r CFU are recovered from Equation (8), when there are n microorganisms on the surface and the probability of an individual microorganism being of a given type is known, is

$$p(r \mid n, \pi_1, \ldots, \pi_K) = \int_{\theta_1=0}^{1} \cdots \int_{\theta_K=0}^{1} p(r \mid n, \pi_1, \ldots, \pi_K, \theta_1, \ldots, \theta_K) \prod_{k=1}^{K} p(\theta_k \mid n, \boldsymbol{r}_k)\, d\theta_1 \cdots d\theta_K. \quad (9)$$

By Bayes' Theorem, the probability that there are n microorganisms on the surface when r CFU are recovered is

$$p(n \mid r) \equiv p(n \mid r, \pi_1, \ldots, \pi_K) \propto p(r \mid n, \pi_1, \ldots, \pi_K) \times p_0(n). \quad (10)$$

If the distribution of r for this assay method is known to be $p(r)$ for surfaces of similar size, composition and microbial population, the unconditional probability of n can be calculated and the probability there are n microorganisms on a surface is $p(n) = \sum_{r=0}^{\infty} p(n \mid r)\, p(r)$. This can be used when a sampling event has not or cannot be performed on the surface.

---

[2] In this study, we assume that π_k are known, fixed quantities for all $k = 1, \ldots, K$. This model readily accommodates the situation where these parameters are unknown or need to be estimated themselves from prior knowledge or data.

[3] The prior on λ is a Jeffreys' Prior. In this case, the Jeffreys' Prior is improper, which leads to an improper prior on n. However, it can be shown that this leads to a proper probability density for all results discussed in this study.

3.4 Comparing Methods for Purposes of Bioburden Estimation

In what follows, we will index each assay method by an integer $q = 1, \ldots, Q$, and refer to assay method q as $\mathcal{M}_q$. We will also index the probability functions above by q to make clear what method is being evaluated. For instance, the probability functions in Equation (10) will be written $p_q(n \mid r)$ to make clear that these probabilities (and all calculations and data going into these probabilities) are relative to method q.

Equation (10) allows us to develop an objective metric to compare the sampling devices and protocols of this study in the context of bioburden estimation. This metric is the expected value of a "cost function". This cost function evaluates the recovery efficiency of a given assay method relative to some reference method, $\mathcal{M}_0$. Ideally, we would have a method $\mathcal{M}^*$ that has a mean probability of 100% that an individual microorganism is recovered, with zero variability — a method that recovers everything from a surface and is perfectly reliable in doing so — and set $\mathcal{M}_0 = \mathcal{M}^*$. In particular, method $\mathcal{M}^*$ implies that the number of microorganisms on a surface sampled, n, equals the number of CFU recovered from the sample, r, with probability one. We measure the cost of deviating from this ideal using the square difference between the modeled bioburden n from method $\mathcal{M}_q$ and the bioburden estimated from the ideal method $\mathcal{M}^*$, $C(\mathcal{M}_q, n \mid \mathcal{M}^*, r) = (n - r)^2$. To compare different methods, we take the expected cost over all possible bioburden estimates, n, when the observed CFU is equal to r; that is, $E_r^*(q) \equiv E[C(\mathcal{M}_q, n \mid \mathcal{M}^*, r)] = \sum_{n=r}^{\infty}(n - r)^2 p_q(n \mid r)$, which can be simplified to reveal how it trades the mean probability that an individual microorganism is recovered with its variability:

$$E_r^*(q) = \sigma_{q,r}^2 + \left(\mu_{q,r} - r\right)^2, \tag{11}$$

where $\mu_{q,r} = \sum_{n=r}^{\infty} n\, p_q(n \mid r)$ is the expected number of CFU on a surface given r have been recovered from method q and $\sigma_{q,r}^2 = \sum_{n=r}^{\infty}\left(n - \mu_{q,r}\right)^2 p_q(n \mid r)$ is the variance of the number of CFU on a surface given r have been recovered from method q. In Equation (11), $\sigma_{q,r}^2$ is directly influenced by the variance in the method's recovery efficiency, penalizing methods with higher variability in their probability of individual CFU recovery. The term $\left(\mu_{q,r} - r\right)^2$ is a direct consequence of the difference between the mean probability of individual CFU recovery of method q with the ideal $\mathcal{M}^*$. This term essentially penalizes a method with a smaller mean recovery efficiency. The cost function tells us how to trade these two — variability and mean recovery efficiency — when comparing the recovery efficiency of different methods.

A particular case of interest we will use in this study is when $r = 0$. This will give us the metric E_0^* to compare assay methods for what is by far the most common case in spacecraft bioburden monitoring, where zero CFU are recovered. Hence, this metric conveys preference for method q over q' when $E_r^*(q) < E_r^*(q')$.

4 RESULTS AND DISCUSSION

The model developed in Section 0 of the experimental process described in Section 2 is applied to analyse recovery efficiency. Recovery efficiency was modelled and analysed holistically with the assay technique deployed, therefore assessments of recovery efficiency are bound to both the sampling device utilized as well as the assay methodology used. Our results for recovery efficiency are often couched in terms of Θ (Theta), the probability that an individual microorganism is recovered. This term is equivalent to the mean recovery efficiency for the reason given in Section 3.2. Theta, however, is used in contexts of bioburden estimation as it is the relevant parameter from a modelling perspective. In this section, we first present results having to do with the sensitivity of recovery efficiency to inoculation level and species, and then summarize overall recovery efficiency results. Finally, we will discuss validation of the mathematical model developed in this study as well as issues of over-dispersion in the data.

4.1 Sensitivity of Recovery Efficiency to Inoculation Level and Species

We address the sensitivity of recovery efficiency to inoculation level by testing if there is a statistically significant trend in the recovery efficiency of *B. atrophaeus* with respect to the targeted inoculation level. As shown in Figure 2, recovery efficiency does not have a strong dependence on inoculation level for swabs. No trend is statistically significant as judged by the 95% credibility interval of the parameter γ_1, although borderline cases exist in cases B and F of the figure. Note that A and E contain no data for the inoculation level of 3. A lower number of replicates were used in C as compared to other experiments, leading to larger scatter in the results relative to other cases.

Figure 3 shows similar graphics for the two wipes assessed by this study. When observations with an inoculation level of 400 CFU are included in the model, there is an evident trend that is statistically significant as judged by the 95% credibility interval of the parameter γ_1 for the TX3211 wipe (A1). However, observations made at this inoculation level are of little practical value to bioburden estimation applications since this result would lead to cleaning of the hardware and resampling prior to bioburden being estimated. Removing these observations from the data results in no trend being statistically significant (A2). The trend observed in A1 indicates that the TX3211 wipe's recovery efficiency may be increase at some inoculation level between 160 and 400 CFU. In contrast, the TX3224 wipe (B1, B2) appears to provide a much more stable, albeit lower recovery efficiency than the TX3211 wipe.

There are significant differences in the recovery efficiency when the sampling device is applied to different species. Figure 4 shows results for the nylon-flocked swab. The reasons for variations in recovery efficiencies are unclear, but it seems possible that different physiochemical adhesive properties, like hydrophobicity or the molecular composition of spore sheaths, can affect the release of spores from surfaces [14]. Other swabs and wipes show similar behaviour in recovery efficiency with respect to species.

4.2 Recovery Efficiency Summary

Table 2 and Table 3 summarize results for the mean recovery efficiency of *B. atrophaeus* and expected cost of each assay method studied, first for swabs, then for wipes. We analyse these groups separately because they are applied in different contexts, with a wipe being applied to much larger surfaces than a swab. Note that while the expected cost presented here captures several important quantitative characteristics of recovery efficiency, it does not consider other factors that are also relevant to making decisions as to which sampling device or protocol to use. For instance, the practicality of using a particular device and assay methodology, handling constraints and accessibility controls in place to ensure the safety of hardware are not considered by this metric. Resources required to revamp processes or change from one protocol to another are also not considered.

For swabs, the nylon-flocked sampling device using the ESA processing technique and facility has the lowest expected cost, and is therefore the swab method preferred by the modelling. This method also has the highest mean recovery efficiency with moderate variability, driving it to have a lower expected cost relative to other methods. A noticeable drop in the mean recovery efficiency occurs with ESA facility nylon-flocked swab experiments when using the NASA (C) instead of ESA (D) standard assay. This result is of borderline statistical significance at the 5% level. Fewer replicates performed for C lead to higher variability in the mean recovery efficiency results. Making the number of replicates consistent with D would allow this difference to be better quantified. The Puritan cotton swab was the second most preferred sampling device. Note there is a borderline statistically significant difference (at the 5% level) between its mean recovery efficiency when tested at a NASA (A) versus ESA (B) facility. Further experiments controlling for processing technique and testing facility would be needed in order to better understand the mechanism causing this difference. Finally, both Copan swabs (E and F) performed worst in these experiments, with mean recovery efficiencies significantly lower and expected costs much higher than other swab assay methods.

Table 2. Mean Recovery Efficiency and Expected Cost for swab sampling devices assessed in this study using *B. atrophaeus*. Values in parenthesis are the endpoints of a 95% credibility interval.

Identifier	Sampling Device	Processing Technique	Testing Facility	Mean Recovery Efficiency	Expected Cost
A	Puritan Cotton Swab	NASA Standard	NASA	31% (26%, 36%)	7
B	Puritan Cotton Swab	NASA Standard	ESA	25% (19%, 31%)	10
C	Nylon Flocked Swab	NASA Standard	ESA	23% (12%, 36%)	17
D	Nylon Flocked Swab	ESA Standard	ESA	38% (32%, 45%)	4
E	Copan Cotton Swab	NASA Standard	ESA	10% (8%, 13%)	60
F	Copan PE Swab	ESA Standard	ESA	10% (3%, 18%)	76

Of the two wipes assessed in this study, the TX3211 wipe has the lowest cost and is therefore the wipe method preferred by the modelling. Despite having much higher variability in its mean recovery efficiency than the TX3224 wipe, the TX3211 wipe has ~2 times the recovery efficiency on average, driving down its expected cost in this comparison.

Table 3. Mean Recovery Efficiency and Expected Cost for wipe sampling devices assessed in this study. Values in parenthesis are the endpoints of a 95% credibility interval.

Identifier	Sampling Device	Processing Technique	Testing Facility	Mean Recovery Efficiency	Expected Cost
A2	TX3211 Wipe	NASA Standard/Milliflex Filter	NASA	27% (6%, 56%)	27
B2	TX3224 Wipe	NASA Standard/Milliflex Filter	NASA	12% (9%, 16%)	46

4.3 Model Validation and Dispersion

The seeding and recovery models developed in Section 3 were validated by assessing how well they predict the observations from the actual experiments performed. Figure 5 shows very good agreement between model and observation. Further validation of the model to assess out-of-sample prediction is on-going.

A primary reason why the model validates well is due to how it captures dispersion in the observations from experiment. In this case, the increase in model complexity from adding the parameter φ is warranted due to several outlier observations that occur in simpler models, leading to over-dispersion. Figure 6 demonstrates this in the case of the TX3211 wipe, but this is representative of other assay methods assessed by this study. Notice how the model in case A of this figure assigns almost zero probability to the observations at 0, 10 and 14 CFU, whereas the model in B does a better job of capturing this variability. Similar results were observed for most other swabs and wipes at most inoculation levels. Note that the model developed in this study captures this dispersion statistically, but does not explain the phenomenon. Further experimentation is required to uncover the mechanism causing this dispersion. Reasons for this dispersion are hypothesized to come from non-uniformities in experiment, such as surface roughness of surfaces sampled. Finally, Figure 7 shows the ramifications of not capturing uncertainty in the seeding model. Treating the number of CFU seeded onto a coupon as a fixed value equal to the average of positive controls can lead to underestimating the variability in the mean recovery efficiency.

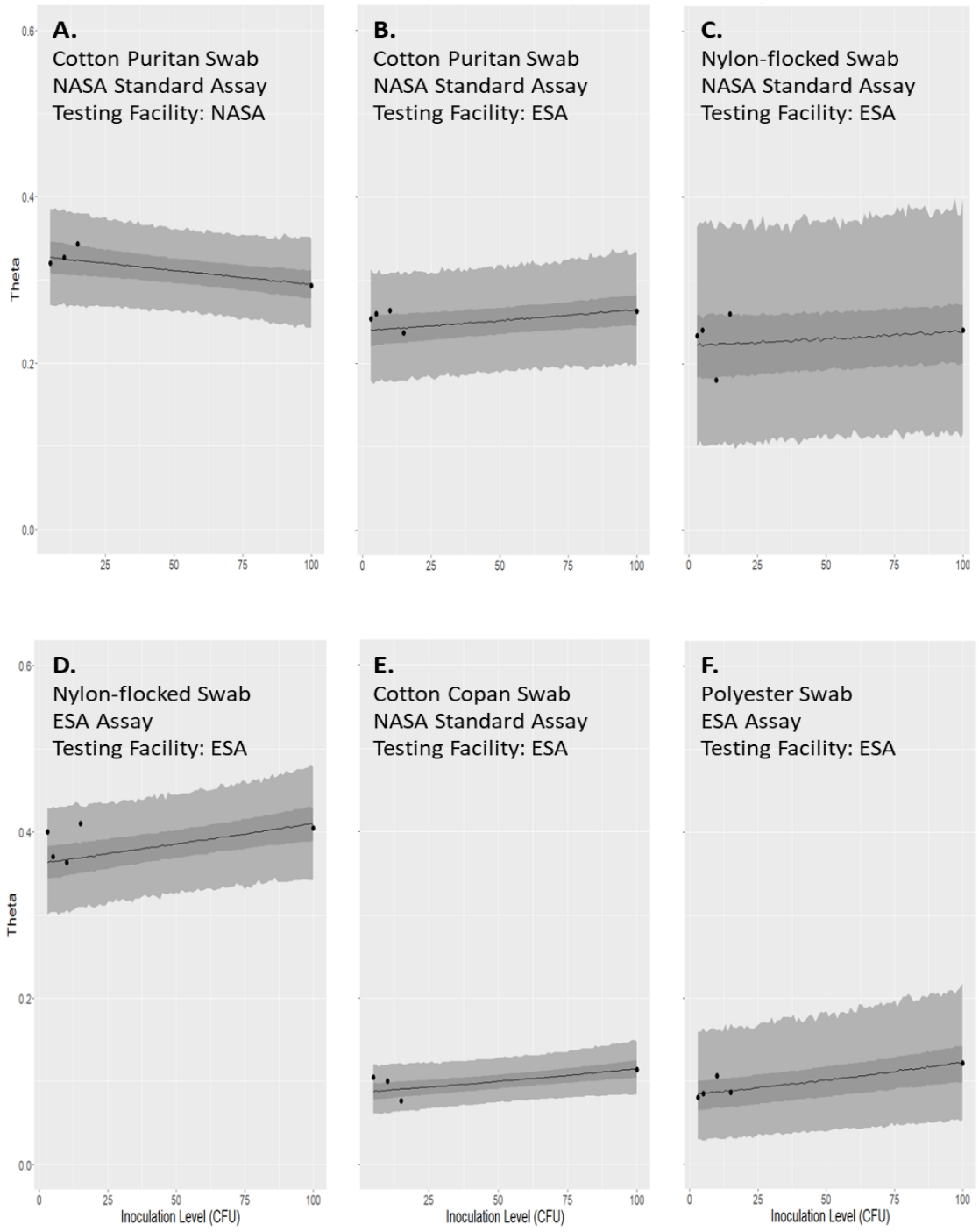

Figure 2. The probability of an individual microorganism being recovered (Theta) with respect to the targeted inoculation level for swab experiments performed in this study with *B. atrophaeus*. The mean value of Theta is shown by the solid line; 50% and 95% credibility intervals for Theta are shown by darker and lighter grey ribbons, respectively (ragged edges of ribbons are due to simulation variation). Calculations of the mean recovery efficiency from experiment are shown by black dots.

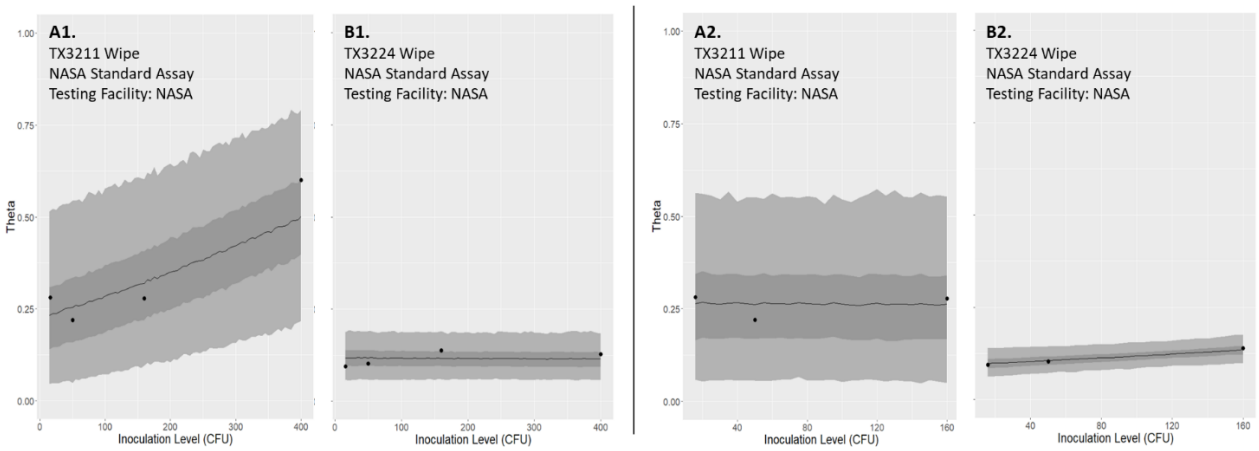

Figure 3. The probability of an individual microorganism being recovered (Theta) with respect to the targeted inoculation level for wipe experiments performed in this study with *B. atrophaeus*. **A1** and **B1** (top row) show Theta when observations from the 400 CFU inoculation level are included in the analysis, while **A2** and **B2** (bottom row) excludes these observations. The mean value of Theta is shown by the solid line; 50% and 95% credibility intervals for Theta are shown by darker and lighter grey ribbons, respectively (ragged edges of ribbons are due to simulation variation). Calculations of the mean recovery efficiency from experiment are shown by black dots.

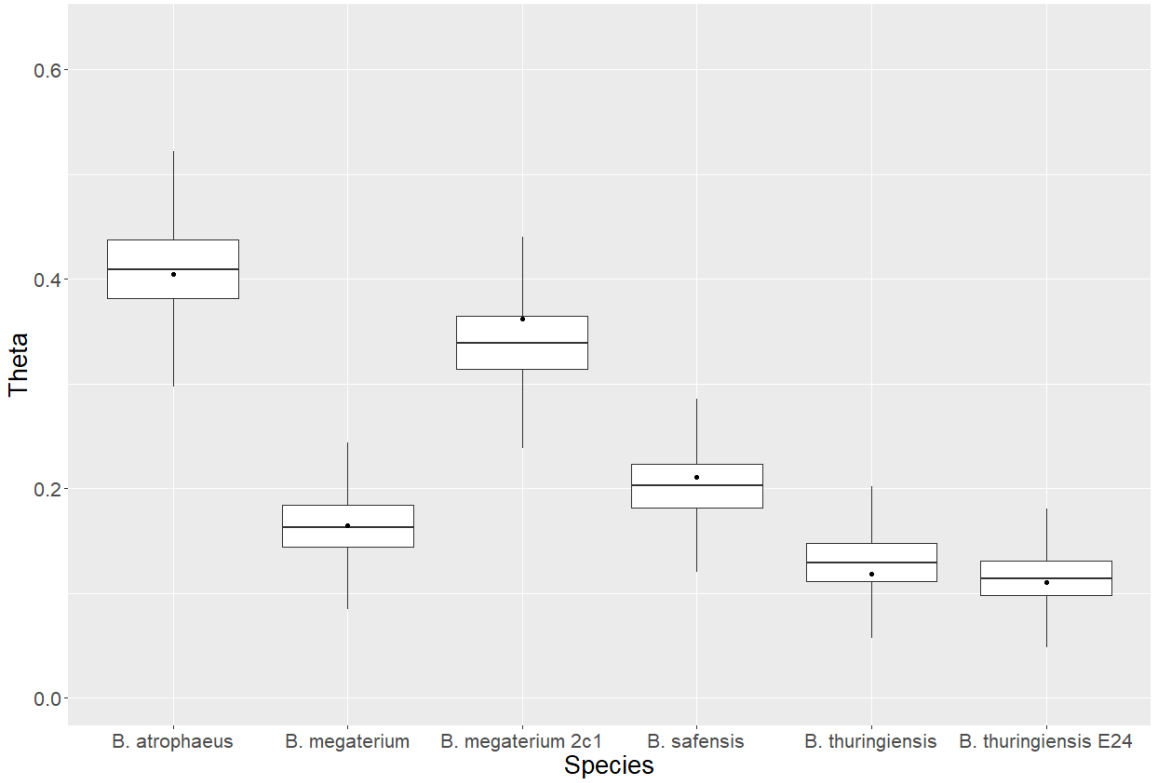

Figure 4. The probability of an individual microorganism being recovered (Theta) with respect to the species for the nylon-flocked swab; target inoculation level = 100 CFU. Box and whisker plots show the modelled middle 50% and 95% ranges, respectively. The model's mean value is shown by the solid horizontal line in the middle of the box. Mean recovery efficiencies calculated from experiment, assuming a fixed inoculant amount equal to the mean of controls, are show by black dots.

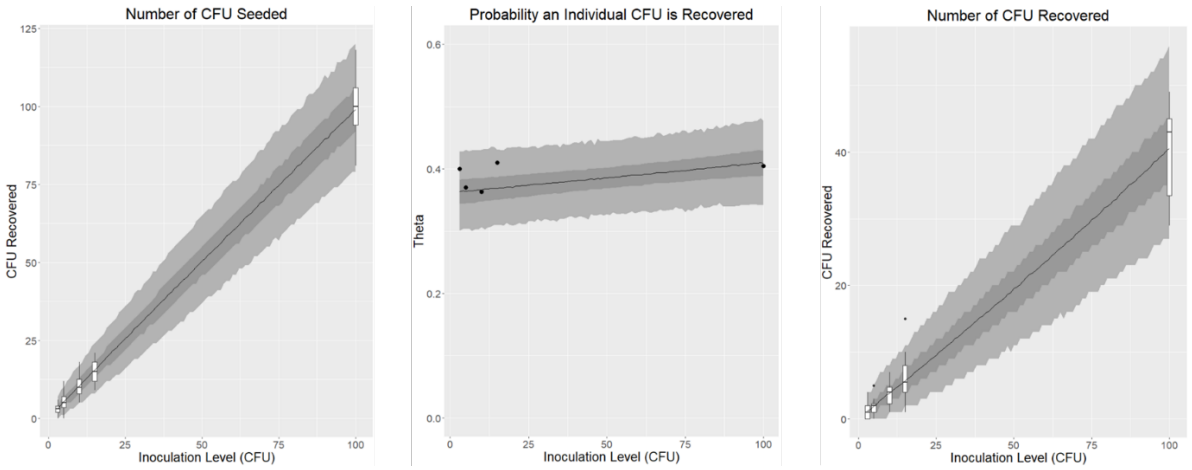

Figure 5. Validation of the mathematical model developed in this study for the nylon-flocked swab (*B. atrophaeus*, ESA facility). On the left, the seeding model predictions (grey ribbons and solid line) are compared with the actual data observations (box-whisker plots) from controls. The center graphic shows Theta, the probability of individual microorganism recovery, with model predictions (grey ribbons and solid line) compared with recovery efficiencies calculated from experiment (black dots). On the right, the integrated seeding and recovery model predictions (grey ribbons and solid line) are compared with the actual data observations (box-whisker plots) from recovery efficiency experiments. The mean value of the model is shown by the solid line; 50% and 95% credibility intervals of the model are shown by darker and lighter grey ribbons, respectively (ragged edges of ribbons are due to simulation variation). Box and whisker plots show the middle 50% and 95% ranges, respectively, calculated from the data. The model validates very well with observation, with the model capturing the dispersion in the observed data.

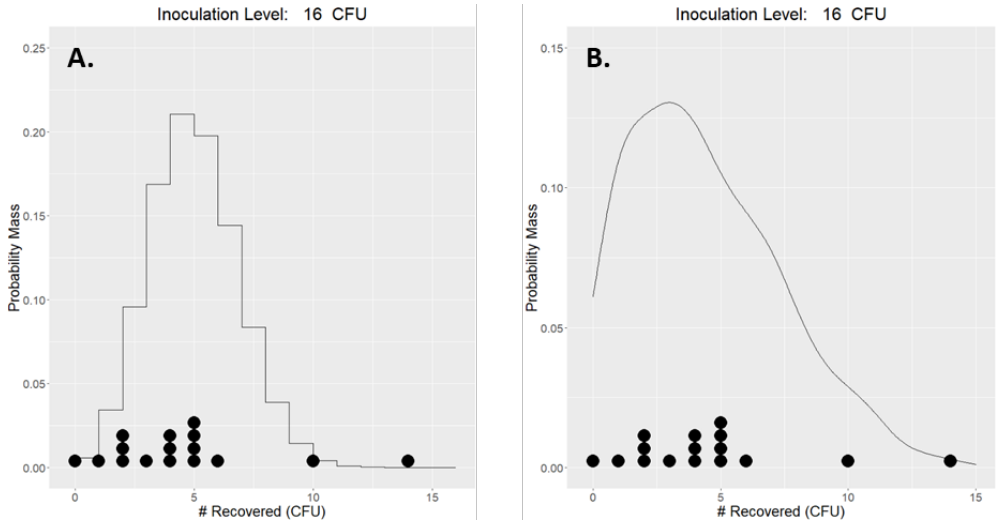

Figure 6. Over-dispersion relative to a simpler model (**A**) warranting a more complex model (**B**) to capture this additional variability in the data. The curves in each plot show model predictions, which is overlaid with a dot plot of the actual observations made from experiment (TX3211 wipe, target inoculation level = 16 CFU).

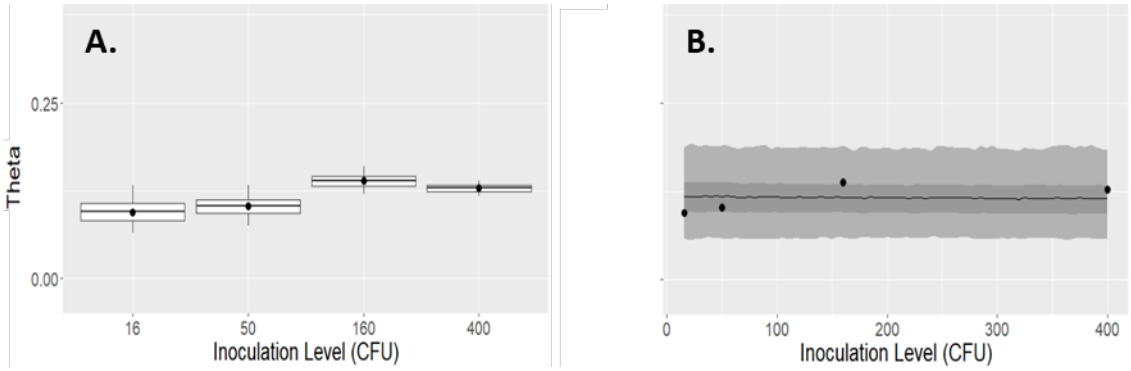

Figure 7. (**A)** The effect of treating the number of CFU seeded onto a coupon as constant (current practice sets this equal to the average of controls) as compared to (**B**) modelling the seeding process as in Section 3.1, for the TX3224 wipe. The method used in **A** does not capture significant variability in the recovery efficiency stemming from uncertainty in the number of CFU seeded onto the coupon. Similar observations hold for other swab and wipe results.

5 CONCLUSION

In this study we have assessed the recovery efficiency of various sampling procedures using a combination of sampling devices and assay processing techniques. A mathematical model was built and informed by controlled experiments, (1) to provide a framework for integration of recovery efficiency to existing Bayesian statistical pipelines used to calculate bioburden density and (2) to determine the ideal sampling device and recovery technique. The nylon flocked swab and TX3211 wipe perform best for purposes of bioburden estimation. For all practical purposes, there does not appear to be a strong trend between recovery efficiency and the inoculation level, although there appears to be an ~2X increase in recovery efficiency of the TX3211 wipe once exceeding a certain inoculation level between 160 and 400 CFU. However, given the stringent bioburden requirements imposed on current Planetary Protection missions <300 - <1000 spores/m^2 CFU counts, a vast majority ~80% of planetary protection CFU counts per sample lie in the range of 0 to 1 CFU. Further characterization of this increase in recovery efficiency may be of value to potential future missions with less stringent requirements. This study shows sensitivity of recovery efficiency to species across all sampling devices, but further experimentation and better knowledge of the distribution of species in cleanrooms is necessary to integrate this variability into bioburden estimation models. The development of molecular based microbial detection techniques may play a critical role in providing a more comprehensive distribution on species present. Additional controlled studies comparing sampling devices (particularly wipes) using both NASA and ESA protocols as well as Milliflex filtration vs standard plating processes may strengthen existing datasets and help improve the mathematical modelling. Finally, the mathematical modelling developed in this study provides the foundation of a rigorous probabilistic tool that can be made available to microbiologists when assessing recovery efficiency and performing bioburden estimation from assays analysed in the lab.

Acknowledgements

The research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration (80NM0018D0004).

1. Puleo JR, Fields ND, Bergstrom SL, Oxborrow GS, Stabekis PD, Koukol R. 1977. Microbiological profiles of the Viking spacecraft. Appl Environ Microbiol 33:379–384.
2. Cooper M, La Duc MT, Probst A, Vaishampayan P, Benardini CS, Piceno YM, Andersen GL, Venkateswaran K. 2011. Comparison of innovative molecular approaches and standard spore assays for assessment of surface cleanliness. Appl Environ Microbiol https://doi.org/10.1128/AEM.00192-11.
3. Hendrickson R, Kazarians G, Benardini JN. 2020. Planetary Protection Implementation on the Interior Exploration Using Seismic Investigations, Geodesy and Heat Transport Mission. Astrobiology https://doi.org/10.1089/ast.2019.2098.
4. Smith SA, Benardini JN, Anderl D, Ford M, Wear E, Schrader M, Schubert W, Deveaux L, Paszczynski A, Childers SE. 2017. Identification and Characterization of Early Mission Phase Microorganisms Residing on the Mars Science Laboratory and Assessment of Their Potential to Survive Mars-like Conditions. Astrobiology 17.
5. Probst A, Facius R, Wirth R, Moissl-Eichinger C. 2010. Validation of a nylon-flocked-swab protocol for efficient recovery of bacterial spores from smooth and rough surfaces. Appl Environ Microbiol 76.
6. Benardini JN, Seuylemezian A, Gribok A. 2020. Predicting Biological Cleanliness: An Empirical Bayes Approach for Spacecraft Bioburden AccountingIEEE Aerospace Conference Proceedings.
7. Gribok A, Benardini JN, Seuylemezian A. 2019. A Bayesian approach for estimating spacecraft microbial bioburden and managing the risk of biological contaminationPSAM 2019 Topical – Practical use of Probabilistic Safety Assessment, 2–3 December 2019, Stockholm, Sweden.
8. Gribok A, Seuylemezian A, Benardini JN. 2020. A Bayesian framework for bioburden density calculations to perform planetary protection probabilistic risk assessment. Proc 30th Eur Saf Reliab Conf 15th Probabilistic Saf Assess Manag Conf.
9. Moeller R, Horneck G, Facius R, Stackebrandt E. 2005. Role of pigmentation in protecting Bacillus sp. endospores against environmental UV radiation. FEMS Microbiol Ecol 51.
10. Stan Development Team. 2020. Stan Modeling Language User's Guide and Reference Manual, Version 2.19.2. Interact Flow Model Lang.
11. RStudio Team. 2021. RStudio: Integrated development environment for R (Version 1.4.1106). J Wildl Manage.
12. Dean Z, DiNicola M, Klonicki E, Roberts S, Clement B, Guan L. Establishing sterility assurance for Bacillus 29669 spores under high heat exposure. Front Microbiol.
13. Cramer JS. 2005. The Origins of Logistic Regression. SSRN Electron J https://doi.org/10.2139/ssrn.360300.
14. Rönner U, Husmark U, Henriksson A. 1990. Adhesion of bacillus spores in relation to hydrophobicity. J Appl Bacteriol 69.