INL/CON-20-57551-Revision-0

Predicting Biological Cleanliness: An Empirical Bayes Approach for Spacecraft Bioburden Accounting

Andrei V Gribok, J. Nick Benardini, Arman Seuylemezian

March 2020

The INL is a U.S. Department of Energy National Laboratory
operated by Battelle Energy Alliance

INL/CON-20-57551-Revision-0

Predicting Biological Cleanliness: An Empirical Bayes Approach for Spacecraft Bioburden Accounting

Andrei V Gribok, J. Nick Benardini, Arman Seuylemezian

March 2020

Idaho National Laboratory
Idaho Falls, Idaho 83415

http://www.inl.gov

Prepared for the
U.S. Department of Energy
Office of Nuclear Energy
Under DOE Idaho Operations Office
Contract DE-AC07-05ID14517

Jet Propulsion Laboratory
California Institute of Technology

Predicting Biological Cleanliness: An Empirical Bayes Approach for Spacecraft Bioburden Accounting

J. Nick Benardini[1], Andrei Gribok[2], Arman Seuylemezian[1]

[1]California Institute of Technology, Jet Propulsion Laboratory

[2]Idaho National Laboratory

2020 IEEE Aerospace Conference
Big Sky Montana
March 9, 2020

Agenda

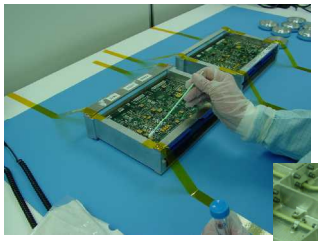
- Biological Cleanliness Verification
- Biological Verification Overview
- Direct Spacecraft Verification
- Direct Verification Count Distribution
- Bioburden Accounting Evolution
- Gamma—Poisson Compound Distribution Model
- Summary of Bioburden InSight Components
- Results and Discussion
- Summary
- Future Work

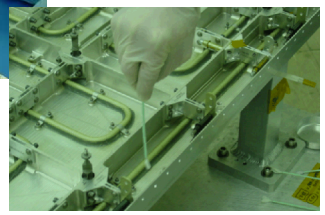
Biological Cleanliness Verification

- How does one calculate spore requirements from observed, direct spacecraft sampling events?
- Directly addresses requirement reporting
  - Mars
    - 5 x 10^5 spores per launched spacecraft
    - 3 x 10^5 spores per landed
    - <300 spores/m2
  - Outer planets
    - Inadvertent contamination of an ocean or other liquid water body to less than a probability of 1 x 10^{-4} per mission.
- NASA or ESA Standard Spore assay for direct hardware verification – damp water wipe or swabs from hardware surfaces during mandatory inspection points.

Biological Verification Overview

Typical Spacecraft Subsystem Process Flow



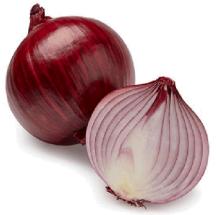


Component

Mandatory Inspection Points

Sub-system

Mandatory Inspection Points



System During I&T

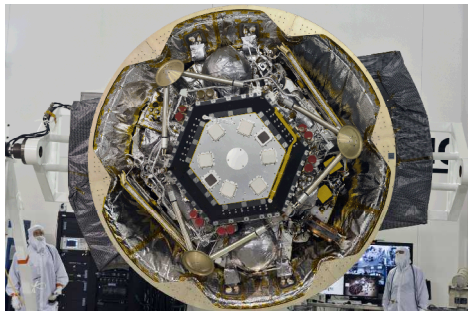


Image Credit:
https://www.flickr.com/people/nasakennedy

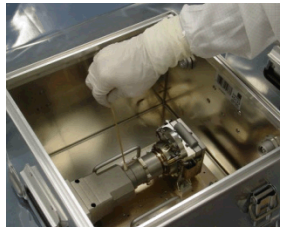
Mandatory Inspection Points

System During Launch Ops

Image Credit:
https://www.flickr.com/people/nasakennedy

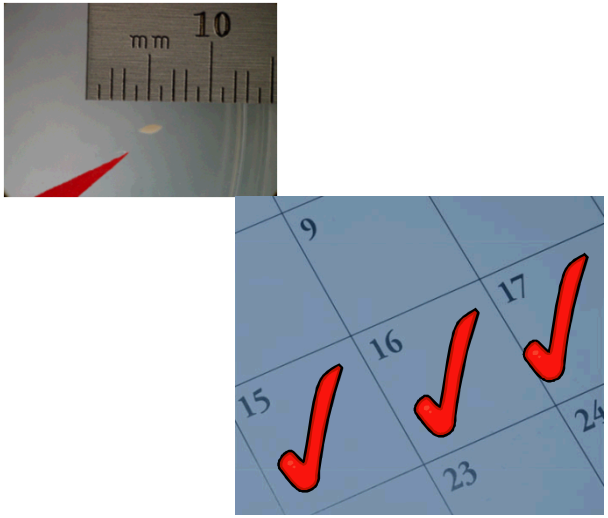
Direct Spacecraft Verification

Clean hardware (and table or bag) with IPA wipe prior to sample or install



Assay Swab or Wipe
(water is used as solvent)

Process swabs and wipes,
~3 hours required post-assay

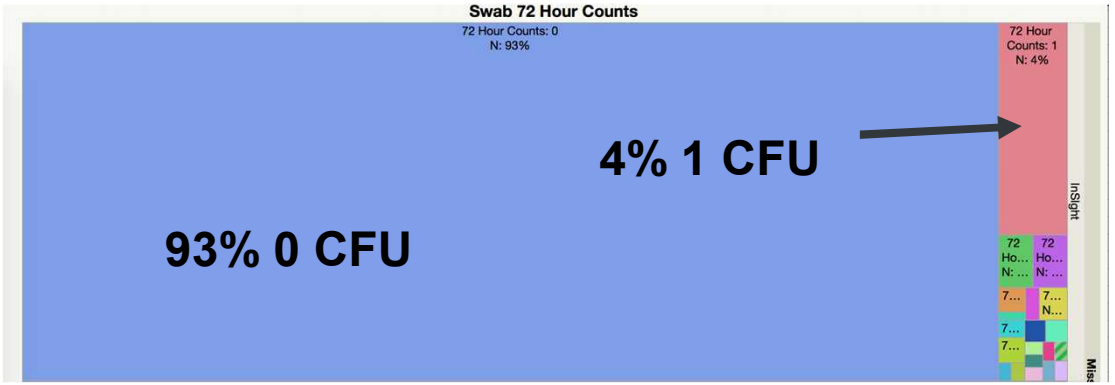


Count Plates for Colony Forming Units (CFU) @ 24h, 48h, and 72h

Direct Verification Counts Distribution

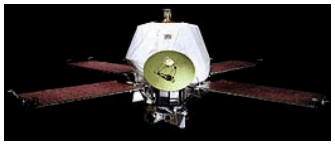
- Bayesian statistics suited dataset similar to that of nuclear industry
  - Data rich with many data points that occur at low frequency
- >80% of wipes and >90% of swabs containing a bioburden count of 0
  - >85% of a missions petri dish observations = no colony forming units

Example of the InSight Missions Dataset n=3,110



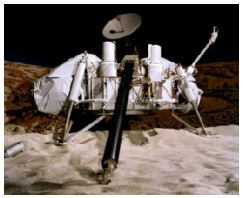
	MER (Both)	MSL	InSight
Swabs	3,066	3,472	1,983
Wipes	529	1,283	1,127
Plates (Equivalent)	25,489	47,997	39,379
Counting Opportunities	76,467	143,991	118,137

Bioburden Accounting for Mars Landed Spacecraft Raw Spore Counts to Calculated Bioburden



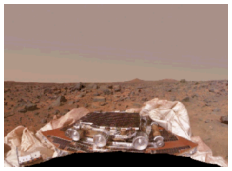
NASA / Artist Illustration

Mariner Mars



NASA / Artist Illustration

Viking Missions



NASA

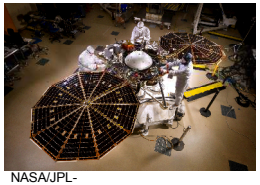
Pathfinder

NASA

Mars Exploration Rovers

NASA

Mars Science Laboratory



NASA/JPL-Caltech/Lockheed Martin

InSight

1970 1980 1990 2000 2010 2020+

Sum of the Means Approach

The total effective area sampled was represented by:

$$A_s = \sum_{i=1}^{n_s} a_{s_i} f_{s_i} e_{s_i}$$

The total number of spores counted was:

$$N_{tot} = \sum_{i=1}^{n_s} N_{s_i}$$

The bioburden density, B, was given by:

$$B = N_{tot} / A_s$$

The estimate of the total bioburden, N, was given by:

$$N = B\, A_0$$

Variable	Definition
A_0	the total area represented by a group or sample set, m^2
n_s, n_w	the total number of swabs or wipes
n_{tot}	the total number of samples
a_{si}, a_{wj}	is the area sampled by the i^{th} swab and the j^{th} wipe, (m^2)
f_s, f_w	the pour fractions for swabs and wipes
e_s, e_w	the recovery efficiencies for swabs and wipes
A_S	the total effective area, (m^2)
N_{si}	the number of CFU counted in the i^{th} swab sample
N_{wj}	the number of CFU counted in the j^{th} wipe sample
N_{tot}	the total number of CFU in a group, (spores/m^2)
B	

Poisson and Gaussian Statistics

Examples equations utilized:

$$\sigma = 1 \div (A_0 A_s)$$

$$B_{max} = 1/A_s + 3\sigma$$

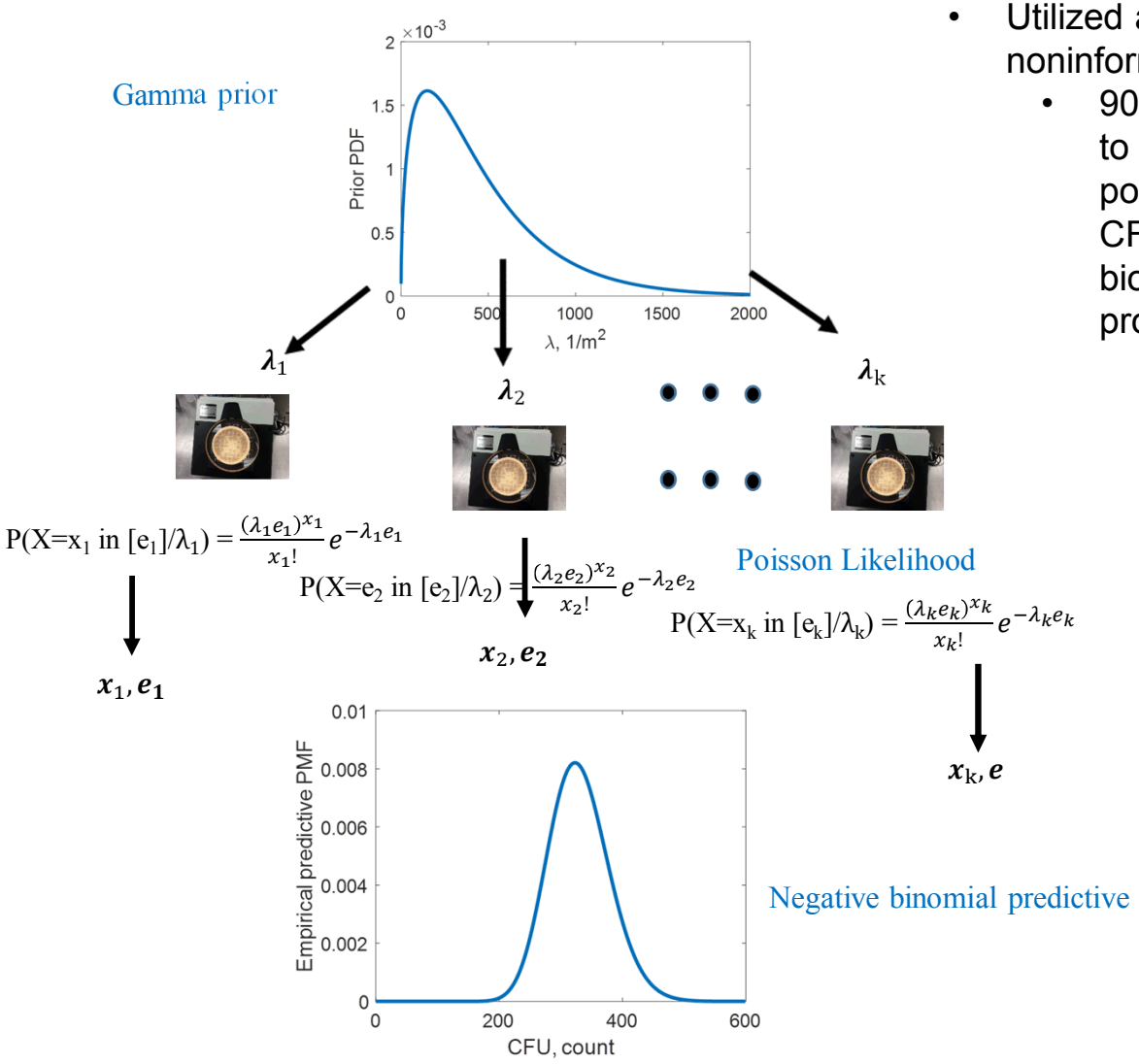
$$\sigma = 1 \div A_s$$

$$B_{max} = 1/A_s + 3\sigma$$

$$\sigma = \sqrt{(N/A_s)}$$

$$B_{max} = N/A_s + 3\sigma$$

Gamma-Poisson Bioburden Compound Distribution Model



- Utilized a developed Constrained noninformative (CNI) prior
  - 90% credible intervals were used to quantify uncertainty in posterior inference. μ=300 CFUs/m^2 reflects the average bioburden density requirement provided

Summary of Bioburden InSight Components

Bounding Cases for Evaluation

Component	CFU count	Area sampled, m^2	Exposure: area sampled × pour ratio, m^2	Total surface area 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

Summary of posterior and predictive inference for components 9, 73, 300

9

Prior distribution	Posterior mean. Bioburden density – λ, CFU/m²	5th percentile of posterior distribution	95th percentile of posterior distribution	Predictive mean, CFU	5th percentile of predictive distribution	95th percentile of predictive distribution
CNI	2.2889	0.0090	8.7928	1.7350	0	7
MOM	0.7603	4.1117e-08	4.0993	0.5763	0	3

73

Prior distribution	Posterior mean. Bioburden density – λ, CFU/m²	5th percentile of posterior distribution	95th percentile of posterior distribution	Predictive mean, CFU	5th percentile of predictive distribution	95th percentile of predictive distribution
CNI	0.8094	0.0031	3.1096	2.2180	0	9
MOM	0.2684	1.4519e-08	1.4475	0.7356	0	4

300

Prior distribution	Posterior mean. Bioburden density – λ, CFU/m²	5th percentile of posterior distribution	95th percentile of posterior distribution	Predictive mean, CFU	5th percentile of predictive distribution	95th percentile of predictive distribution
CNI	2.2315	0.2617	5.8130	11.1579	1	30
MOM	1.7355	0.1267	4.9272	8.6778	0	26

Summary of posterior and predictive inference for components 169, 283, 243

169

Prior distribution	Posterior mean. Bioburden density – λ, CFU/m²	5th percentile of posterior distribution	95th percentile of posterior distribution	Predictive mean, CFU	5th percentile of predictive distribution	95th percentile of predictive distribution
CNI	7.7452	0.9083	20.1757	4.5309	0	13
MOM	6.0352	0.4408	17.1337	3.5305	0	11

283

Prior distribution	Posterior mean. Bioburden density – λ, CFU/m²	5th percentile of posterior distribution	95th percentile of posterior distribution	Predictive mean, CFU	5th percentile of predictive distribution	95th percentile of predictive distribution
CNI	4.8059	1.9987	8.5961	57.6713	22	105
MOM	4.5158	1.8133	8.2011	54.1903	20	100

243

Prior distribution	Posterior mean. Bioburden density – λ, CFU/m²	5th percentile of posterior distribution	95th percentile of posterior distribution	Predictive mean, CFU	5th percentile of predictive distribution	95th percentile of predictive distribution
CNI	47.5504	19.7758	85.0510	14.1700	4	27
MOM	44.8607	18.0138	81.4700	13.3685	4	26

Summary of posterior and predictive inference for components 38, 261

38

Prior distribution	Posterior mean. Bioburden density – λ, CFU/m²	5th percentile of posterior distribution	95th percentile of posterior distribution	Predictive mean, CFU	5th percentile of predictive distribution	95th percentile of predictive distribution
CNI	15.4671	9.0398	23.2949	154.6710	88	236
MOM	15.0630	8.7294	22.7980	150.6308	85	231

261

Prior distribution	Posterior mean. Bioburden density – λ, CFU/m²	5th percentile of posterior distribution	95th percentile of posterior distribution	Predictive mean, CFU	5th percentile of predictive distribution	95th percentile of predictive distribution
CNI	1057.0469	829.0645	1307.8988	329.7986	253	414
MOM	1061.3672	831.7609	1314.0833	331.1465	254	416

Bioburden Density Comparison

	Proposed Bayesian Approach		MSL-based 3 sigma (NASA Legacy)	InSight-based weighted average technique (NASA Current)
Component	CNI, Posterior Mean Bioburden Density – λ, CFU/m^2	MOM, Posterior Mean Bioburden Density – λ, CFU/m2	3 sigma Bioburden Density – λ, CFU/m^2	Weighted Average Bioburden Density – λ, CFU/m^2
9	2.2889	0.7603	13.84	27.99
73	0.8094	0.2684	4.87	17.36
300	2.2315	1.7355	5.96	9.54
169	7.7452	6.0352	20.83	33.70
283	4.8059	4.5158	5.17	11.11
243	47.5504	44.8607	130.14	186.70
38	15.4671	15.0630	52.06	9.66
261	1057.0469	1061.3672	2349.53	658.47

Summary

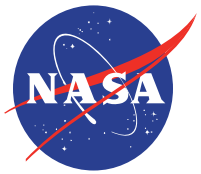
- The mean values of predictive distribution are correlated with the total CFU count found on the component
- In general, the uncertainty in the posterior estimate of the bioburden density depends on the number of counts and sampled area, while the uncertainty in the predictive inference depends on sampling percentage.
- Bayes Factor approach is indicative of reliance on the data rather than on the prior for the components with larger counts
- Implementing a Bayesian statistical approach to perform bioburden density estimations will:
  1. facilitate the application of historical datasets and engineering judgement in estimating the total bioburden and bioburden density;
  2. assign appropriate confidence intervals and account for uncertainty using a methodological approach; and
  3. allow for the prediction of bioburden throughout the lifecycle of a project.
- Future work to include
  - Hierarchical Bayes being the subject of future work and model validation
  - Develop informed priors based on hardware and reporting use cases
  - Expansion of statistical approach to account for complete bioburden reporting of direct, implied and specification value hardware groups.

Abstract

To comply with the international planetary protection policy set forth by the Committee on Space Research and NASA Agency level requirements, spacecraft destined to biologically sensitive planetary bodies have to minimize terrestrial biological contamination. Analysis, testing and inspection are the standard forward verification activities that are used to demonstrate compliance with the biological contamination requirements. For testing of spacecraft surface areas, a swab or wipe sample is collected from surfaces prior to last access and subsequently processed in the lab using NASA Approved Planetary Protection Methods for Culture Based Assays. Raw data resulting from this assay is then statistically treated employing a mathematical paradigm stemming from the 1970's Viking Lander Project to generate the bioburden density and total microbial bioburden present. This standard approach arbitrarily accounts for error and provides an upper conservative bound as it reports the maximum number of spores estimated to be present on flight hardware surfaces. A bioburden density estimate factors in the following variables: the observed bioburden count, representative volume processed, sampling efficiencies. Notably, to account for error in the approach, a 0 observed count is arbitrarily changed to a count of 1 for each hardware grouping.

The data generated by spacecraft bioburden verification campaigns in the past have resulted in <80% of wipes and <90% of swabs containing a bioburden count of 0. As such, having a robust and well documented statistical approach for dealing with the probability of low incident rates is necessary to be able to estimate spacecraft bioburden. Being able to statistically describe the bioburden distribution and associated confidence level is a gamechanger for the development of bioburden allocations during mission design and will allow for tighter management of risk throughout spacecraft build. Thus, Empirical Bayes statistical approach was evaluated to estimate the microbial bioburden on spacecraft to mitigate the aforementioned mathematical concerns and provide a probabilistic bioburden distribution of the flight hardware surface.

For application of this approach to performing bioburden calculations, a range of non-informative prior assumptions on hardware surfaces are explored for Bayesian analyses while informative priors using posterior distributions from prior assays are utilized for Empirical Bayes analyses. Several non-informative priors are currently under investigation to assess fitness including use of these priors to serve as a foundation to build off of NASA specification values or a basis of risk to account for unknowns during the integration and testing process. Informative priors under consideration are generated using sampled bioburden values from hardware originating within like processing environments (e.g. vendor cleaning process or similar assembly process), temporal spacecraft status events as a prediction for hardware cleanliness of future samples, and heritage system bioburden actuals to predict allocation for subsequent missions. Informative priors and probabilistic bioburden distributions are then validated using data sets from the Mars Exploration Rover, Mars Science Laboratory, and InSight missions. Using Empirical Bayes approach to generate a probabilistic bioburden distribution as demonstrated through mission use cases provides a valid approach for use in the end-to-end requirements verification process.



Jet Propulsion Laboratory
California Institute of Technology

jpl.nasa.gov