50th International Conference on Environmental Systems
12-16 July 2020, Lisbon, Portugal

ICES-2020-314

More Data Needed for Failure Rate Estimation, Validation, and Uncertainty Reduction

Andrew C. Owens,[1] Nicole E. Piontek,[2] and William M. Cirillo[3]
NASA Langley Research Center, Hampton, VA 23681, USA

Chel Stromgren[4] and Jason Cho[5]
Binera, Inc., Rockville, MD 20850, USA

Current Environmental Control and Life Support System (ECLSS) development and test activities are not generating data fast enough to provide statistically-supportable precise Orbital Replacement Unit (ORU) failure rate estimates for future missions. Accurate and precise failure rate estimates are critical for missions beyond Low Earth Orbit (LEO) because current risk mitigation approaches – namely regular resupply and rapid abort capabilities – will not be available. Safe operations will depend on mission planners' ability to accurately forecast spares demand and efficiently provide the necessary resources. However, even after more than a decade of operations on board the International Space Station (ISS), a significant amount of uncertainty remains in failure rate estimates. Uncertain or inaccurate failure rates result in increased risk and spares mass for future missions. A Bayesian failure rate estimation approach, such as the one currently implemented by the ISS Program, can help reduce uncertainty by incorporating engineering judgement into failure rate estimates. However, experience on the ISS and with other complex systems shows that these prior failure rate estimates are often inaccurate. In addition, prior failure rate estimates are typically point values; some level of uncertainty must be added to convert these into probability distributions for Bayesian updating, and there are several potential methods for doing so. Due to the low rate of data collection, these subjective (and often inaccurate) prior estimates currently have a strong influence on the end result. This paper examines the challenges associated with failure rate estimation, validation, and uncertainty reduction in the context of ECLSS development for beyond-LEO missions. A variety of techniques for generating and updating Bayesian priors are discussed and evaluated using both real-world and simulated data. Potential solutions for improving failure rate estimation, including testing additional units, are analyzed and discussed, and a set of recommendations are provided for next-generation system development activities.

Nomenclature

α	=	Gamma shape parameter
β	=	Gamma scale parameter
γ	=	Variance-to-Expectation ratio
ε	=	Error factor
Λ	=	Failure rate estimate (random variable)
λ	=	Failure rate (point value)
$\bar{\lambda}$	=	Expected failure rate
σ^2	=	Variance
n_o	=	Number of observed failures

[1] Aerospace Engineer, Space Mission Analysis Branch, 1 N. Dryden St. MS 462
[2] Intern, Space Mission Analysis Branch, 1 N. Dryden St. MS 462
[3] Senior Researcher, Space Mission Analysis Branch, 1 N. Dryden St. MS 462
[4] Vice President / Chief Scientist, 77 S. Washington St.
[5] Aerospace Engineer, 77 S. Washington St.

t_o	=	Total accumulated operating time
ECLSS		Environmental Control and Life Support Systems
ISS		International Space Station
LEO		Low Earth Orbit
MTBF		Mean Time Between Failures
ORU		Orbital Replacement Unit
PDF		Probability Density Function
POS		Probability of Sufficiency
V2E		Variance-to-Expectation

I. Introduction

FUTURE crewed missions beyond Low Earth Orbit (LEO) will be logistically isolated for longer than ever before. Without access to the abort and resupply options that are used to mitigate risk and manage supplies for the International Space Station (ISS) and other human spaceflight missions, crews will have to maintain systems and perform repairs using only the supplies they carry with them. Spare parts in particular will be a significant driver of mass and risk. Unlike other logistical supplies with deterministic (or nearly deterministic) demand, such as water or oxygen, demand for spare parts is driven by random failure processes. There is no way to predict exactly how many spares will be required for a given mission. Instead, the probability distribution describing the number of spares that may be needed during the mission must be generated based on failure rate estimates for each Orbital Replacement Unit (ORU), and then spares are allocated to achieve a risk target, usually defined in terms of Probability of Sufficiency (POS). As a result, the risk and mass associated with spares for a given mission is dependent on the failure rates of the ORUs within the system.[1–3]

Unfortunately, an ORU's failure rate is not a parameter that can be measured directly, since it is not a physical characteristic of the ORU but rather a parameter describing the distribution of the amount of time that the ORU will operate before random failure. Instead, failure rates must be estimated based on analogy to previous systems and/or statistical analysis of operational or test data. As a result, failure rates are inherently uncertain estimates rather than specific values, and this uncertainty presents significant challenges for future mission supportability.[4,5] This uncertainty can be reduced by testing and gathering more data, but significant amounts of data are required to generate precise statistical estimates of failure rates. Even after more than a decade of in-space operations aboard the ISS, a significant amount of uncertainty remains in failure rate estimates for Environmental Control and Life Support (ECLSS) ORUs, and this uncertainty results in higher spares mass requirements for future missions.[6,7]

A common approach for managing failure rate uncertainty is to apply Bayesian updating to a probability distribution describing an initial failure rate estimate in order to generate an updated failure rate distribution that more closely reflects observed system behavior.[8] However, prior failure rate estimates are typically only available as point values. Therefore, some method must be applied to set the initial amount of uncertainty in the estimate and transform the point value in to a probability distribution to be updated. In addition, it is very difficult to generate accurate prior estimates for new systems,[9] and ISS experience has shown that prior failure rate estimates are often inaccurate, sometimes by a significant amount.[4,7]

This paper describes and evaluates four techniques for initializing the level of uncertainty in prior failure rate estimates for Bayesian updating, including applying an equal error factor, equal Variance-to-Expectation (V2E) ratio, or equal variance to all estimates as well as the Jeffreys noninformative prior. Section III applies each of these methods to simulated test results from an item with a known failure rate in order to evaluate the accuracy of resulting estimates and assess sensitivities to inaccurate prior estimates. Section IV then examines the impact of different prior initialization techniques on spares estimation for a notional Mars mission, in terms of both total spares mass and the number of spares provided for each ORU. Sections V and VI present discussion and conclusions based on these results.

II. Methodology

A. Bayesian Failure Rate Estimation

Bayesian failure rate estimation is the process of using data to update an initial failure rate estimate, called a prior, to generate an improved estimate, called a posterior. Both the prior and posterior failure rate estimates are random variables – denoted Λ_{pri} and Λ_{post}, respectively – whose probability distributions describe the belief in the value of the true failure rate. For a process with an assumed constant failure rate (a common assumption in reliability and supportability analysis[10]), failure rate estimates are typically assumed to follow a lognormal or gamma distribution. This research uses the gamma distribution, since it is a conjugate prior for a Poisson process and therefore enables closed-form Bayesian updating.[8] One of the two Bayesian failure rate updating processes used by the ISS program uses the gamma distribution. A gamma distribution is parameterized by a shape parameter α and scale parameter β, which are related to the expected value and variance of the estimate by the following equations:[8,11]

$$\alpha = \frac{\mathrm{E}[\Lambda]^2}{\mathrm{Var}[\Lambda]} \tag{1}$$

$$\beta = \frac{\mathrm{E}[\Lambda]}{\mathrm{Var}[\Lambda]} \tag{2}$$

Bayesian updating calculates posterior parameters by updating prior parameters using total accumulated operating time t_o and the number of failures observed during that time n_o.[8]

$$\alpha_{post} = \alpha_{pri} + n_o \tag{3}$$

$$\beta_{post} = \beta_{pri} + t_o \tag{4}$$

The posterior parameters can then be used to calculate the resulting expected value, variance, median, or any other value of interest from the posterior failure rate estimate Λ_{post}. When combined with other ORU information, they can also be used for spares allocation optimization and POS/spares mass estimation for a particular system and mission.[3]

B. Prior Uncertainty Initialization Methodologies

As noted in Section I, the engineering estimates of ORU failure rates that form the basis of initial failure rate estimates are typically only available as point estimates. Some level of uncertainty must be incorporated into these estimates to produce the failure rate distributions required for Bayesian analysis. While the most effective method would be to generate prior distributions individually for each ORU instead of just point estimates, this approach would likely require a significant amount of additional effort. A more common approach is to set the uncertainty in the prior estimate as a function of the initial failure rate, thereby allowing that single parameter to define a prior distribution. The following subsections describe three methods for generating prior distributions from point estimates. In each case, the initial point estimate is used as the expected failure rate, $\bar{\lambda}$, and that value along with some other parameter are used to define the gamma parameters used for Bayesian updating. A fourth method is also described that does not use the initial point estimate at all, but instead uses a standard noninformative prior, which is a method for attempting to initialize a Bayesian estimate without prior bias.

1. Equal Error Factor

A common approach for space system failure rate estimation is to apply the same error factor to all ORUs. For example, this is the approach used by the ISS program to estimate ORU failure rates for ISS systems. The error factor is defined as the ratio of the 95th percentile of the failure rate distribution, and effectively defines the width of the uncertainty distribution in conjunction with the expected failure rate. This parameterization is often used in conjunction with a lognormal model to calculate expected value and variance, which can then be used to form a gamma approximation for Bayesian updates. Given an expected failure rate $\bar{\lambda}$ and error factor ε, the variance of the lognormal failure rate distribution are:[11]

$$\mathrm{Var}[\Lambda] = \bar{\lambda}^2 \left(e^{\left(\frac{\ln \varepsilon}{1.645}\right)^2} - 1 \right) \tag{5}$$

Equation 5 shows that applying the same error factor to all ORUs is equivalent to specifying that the amount of variance in the estimate is proportional to the square of the failure rate. When used with equations 1 and 2, this enables the direct calculation of gamma parameters given $\bar{\lambda}$ and ε:

$$\alpha = \frac{\bar{\lambda}^2}{\bar{\lambda}^2 \left(e^{\left(\frac{\ln \varepsilon}{1.645}\right)^2} - 1 \right)} = \frac{1}{e^{\left(\frac{\ln \varepsilon}{1.645}\right)^2} - 1} \tag{6}$$

$$\beta = \frac{\bar{\lambda}}{\bar{\lambda}^2\left(e^{\left(\frac{\ln \varepsilon}{1.645}\right)^2} - 1\right)} = \frac{1}{\bar{\lambda}\left(e^{\left(\frac{\ln \varepsilon}{1.645}\right)^2} - 1\right)} \tag{7}$$

Note that in this case α depends only on the error factor ε, and therefore this approach is equivalent to specifying that all prior failure rate estimates have the same shape parameter α.

2. Equal Variance-to-Expectation (V2E) Ratio

Another potential approach to prior initialization is to specify that the amount of variance in the estimate is proportional to the expected failure rate. Given a V2E ratio γ, defined as the variance of the distribution divided by the expected value, the variance of an estimate is

$$\mathrm{Var}[\Lambda] = \gamma\bar{\lambda} \tag{8}$$

The gamma parameters are then

$$\alpha = \frac{\bar{\lambda}}{\gamma} \tag{9}$$

$$\beta = \frac{1}{\gamma} \tag{10}$$

Note that this approach is equivalent to specifying that all prior estimates use the same scale parameter β.

3. Equal Variance

The final approach examined here that makes use of the initial failure rate estimate simply specifies that all prior estimates have the same variance, denoted σ^2. In this case, the gamma parameters are

$$\alpha = \frac{\bar{\lambda}^2}{\sigma^2} \tag{11}$$

$$\beta = \frac{\bar{\lambda}}{\sigma^2} \tag{12}$$

Under this approach, both gamma parameters depend on the initial failure rate estimate.

4. Jeffreys Noninformative Prior

The fourth approach examined in this paper does not use the initial failure rate estimate, but instead uses the Jeffreys noninformative prior, a standard approach for generating a Bayesian prior that contains very little initial information. Effectively, the Jeffreys prior is an approach that provides a starting point for Bayesian updating that doesn't inject much bias into the process; put another way, Bayesian updates that use the Jeffreys prior are (in theory) driven more by the observed data than by prior beliefs. For a Poisson process, the standard Jeffreys prior is a gamma distribution with $\alpha_{pri} = 0.5$ and $\beta_{pri} = 0$. This is considered a conservative approach for failure rate estimation, since it effectively initializes the estimate with a "half failure" at time $t_o = 0$.[8] While the noninformative prior approach has the benefit of avoiding the risk of incorrect prior failure rate estimates, it also typically results in a greater amount of uncertainty in the estimate, since it does not make use of information provided by prior estimates.

C. Spares Mass and POS Estimation

This paper applies the spares mass and POS model described by Owens,[3] which has been applied and described in detail in several previous papers.[5–7] Given a set of ORUs with associated supportability data – including mass, quantity, duty cycle, and failure rate estimate – the model assesses the POS for each ORU as a function of the number of spares provided for that ORU using a negative binomial distribution, which is equivalent to the distribution of the number of events resulting from a Poisson process with an uncertain, gamma-distributed parameter.[11,12] The spares mass required for a given mission endurance and POS requirement is determined by optimizing the spares allocation to achieve that POS requirement while minimizing total spares mass.[3,13–15]

III. Impact of Initialization Method on Failure Rate Estimates

This section explores the differences between the different methods described in Section II.B on posterior failure rate estimates for a notional ORU with a notional test result as a function of the initial failure rate estimate.

A. Updates from Simulated Test Results

Assume 87,600h (10yrs) of continuous testing were conducted. This could represent a dedicated ground test of this particular ORU, or it could represent on-orbit operations on the ISS. Each time the test unit failed, it was immediately replaced with an identical unit and the test continued. In order to keep this example simple, assume that only one unit was operated at a time; in practice, test data gathering can be accelerated by operating multiple units in parallel. Test results were simulated using a Poisson process with an assumed failure rate of 4.57×10^{-5}h, which corresponds to a Mean Time Between Failures (MTBF) of 21,900h (2.5yrs). In total, 4 failures occurred, with the failed units having accumulated 15,869h, 50,799h, 9,066h, and 10,145h of operation, respectively. The 5th unit, which was still operational at the end of the test, had accumulated 1,719h of operation. It is important to note that these simulated results represent just one many possible outcomes that could have occurred, even with the failure rate fixed at one value. The amount of time that a particular ORU will operate before failure is random, and if this simulation were repeated it would generate a different set of failure times, potentially resulting in more or fewer failures during the 10-year test period.

In practice, test results cannot be known before the test is executed, and failure rate estimates made before the test do not have the benefit of those results. Historical experience has shown that prior failure rate estimates are often over- or underestimated, sometimes significantly.[4,7] In order to evaluate the prior initialization methodologies described above relative to each other, this paper examines three cases:

- *Accurate prior*: the initial failure rate estimate is 4.57×10^{-5}h, which was the value used to generate the simulated test results and represents the "true" value of this simulated case.
- *Underestimate*: the initial failure rate is estimate is 9.13×10^{-6}h, which is lower than the "true" value by a factor of 5. This represents a case where the failure rate was underestimated, and the ORU turned out to be less reliable than expected.
- *Overestimate*: the initial failure rate estimate is 2.28×10^{-4}h, which is higher than the "true" value by a factor of 5. This represents a case where the failure rate was overestimated, and the ORU turned out to be more reliable than expected.

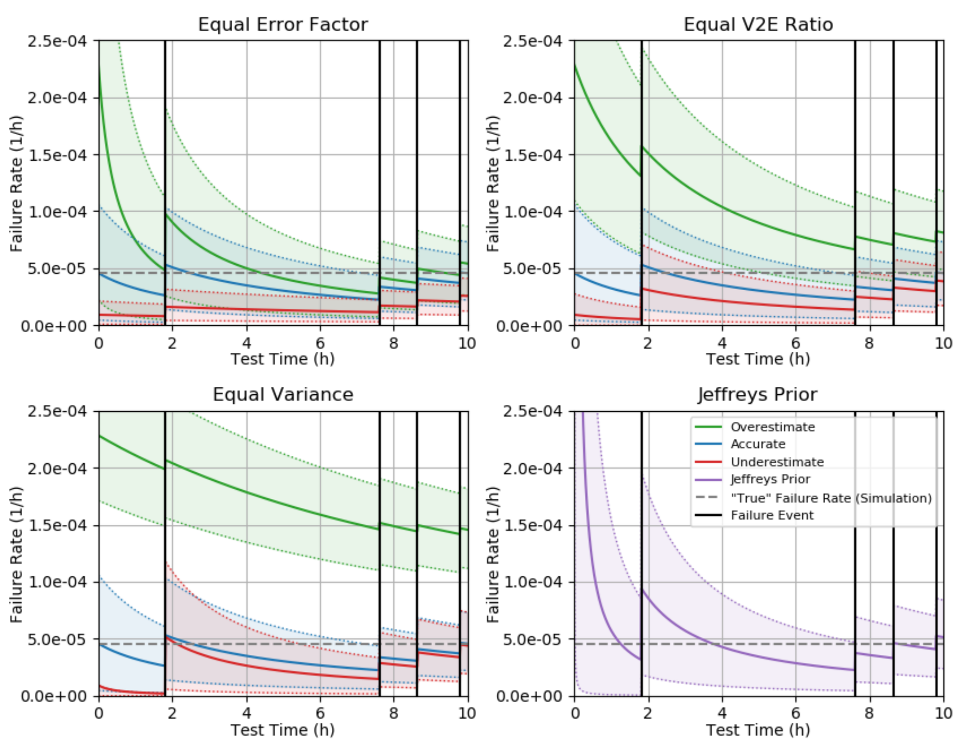

Figure 1: Timeline of results from Bayesian updating based on simulated test data for each prior initialization technique for accurate priors as well as over- and underestimates.

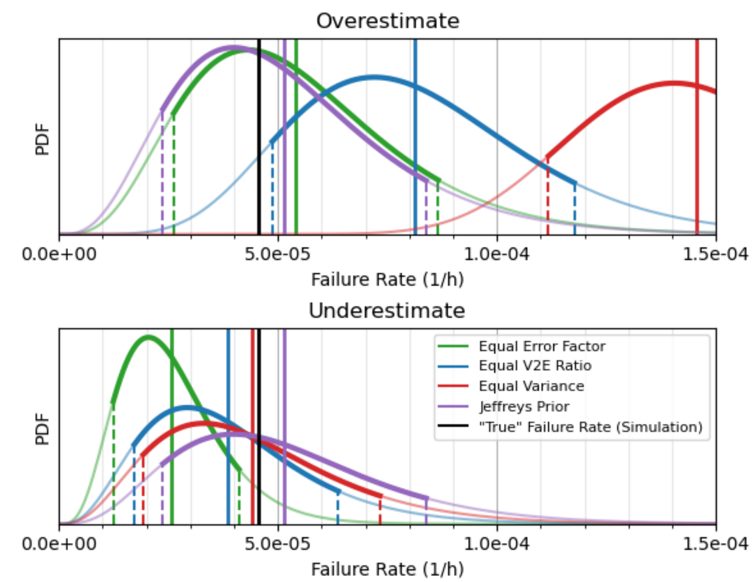

Figure 2: Failure rate estimate PDFs after Bayesian updating when the prior is an overestimate (top) or underestimate (bottom). Solid vertical lines indicate the expected failure rate after updating, and the bolded section of each curve shows the 80% credible interval.

For each case, the expected value and 80% credible interval of the failure rate estimate is plotted over the test timeline for each prior initialization technique. For the purposes of this comparison, the initial error factor ε is assumed to be 4, a common value for space system Bayesian reliability estimation. Since the primary interest in this investigation is how the Bayesian estimate responds to inaccurate priors, the V2E ratio γ and variance σ^2 are set to values that produce the same variance for the accurate prior case; that is, the prior failure rate estimate is the same for all initialization methodologies when the initial failure rate estimate is accurate. With an error factor of 4 and an initial failure rate $\bar{\lambda}$ of 4.57×10^{-5}h^{-1}, equation 5 yields a prior variance of 2.16×10^{-9}h^{-2}. Thus the equal variance parameter σ^2 is set to 2.16×10^{-9}h^{-2} and the equal V2E ratio parameter γ is set to 4.72×10^{-5}h^{-1}.

Figure 1 shows the evolution of failure rate estimates over time for the simulated test described above, using overestimated, accurate, and underestimated initial failure rates as well as different prior initialization methods (shown in different quadrants of the figure). Results using the Jeffreys prior (purple) are also shown. The so-called "true" failure rate, which was used to simulate test results, is indicated by the gray horizontal dashed line, and failure events are indicated by vertical black lines. Figure 2 shows the Probability Density Function (PDF) of the failure rate estimate for the over- and underestimate cases, along with vertical lines indicating the expected failure rate after updating and the 80% credible interval. Note that the colors in Figure 2 do not correspond to those in Figure 1, since they are used to indicate estimation technique instead of over-/underestimation. In addition, the upper bound on the credible interval for the Equal Variance estimate when the initial failure rate is overestimated, which was very high, was cut off in order to focus the x-axis range on the area of interest. These results show that the results of Bayesian updating can be significantly different when the prior has a different initial amount of variance, even if the same data are supplied to the update. In all cases, as expected, the estimate approaches the "true" failure rate value as more data are accumulated, and the level of uncertainty in the estimate (indicated by the width of the 80% credible interval) decreases. However, depending on the level of inaccuracy in the initial failure rate estimate and the method used to initialize uncertainty, this updating process could require years of operating experience,

Table 1: Summary of results after 10 years of simulated testing as a function of different prior accuracies and initialization methods. For this notional example the "true" failure rate used to generate simulated test results is 4.57×10^{-5}h^{-1}, and the over- and underestimated priors were off by a factor of 5.

		Equal Error Factor	Equal V2E Ratio	Equal Variance	Jeffreys Prior
Over	Expected Failure Rate (h^{-1})	5.41×10^{-5}	8.12×10^{-5}	1.46×10^{-4}	5.14×10^{-5}
	% Error	+18.4%	+77.9%	+218.9%	+12.5%
	"True" Value in 80% Credible Interval?	Yes	No	No	Yes
	Variance in Failure Rate (h^{-2})	5.89×10^{-10}	7.47×10^{-10}	7.53×10^{-10}	5.86×10^{-10}
Accurate	Expected Failure Rate (h^{-1})	4.57×10^{-5}	4.57×10^{-5}	4.57×10^{-5}	5.14×10^{-5}
	% Error	0.0%	0.0%	0.0%	+12.5%
	"True" Value in 80% Credible Interval?	Yes	Yes	Yes	Yes
	Variance in Failure Rate (h^{-2})	4.20×10^{-10}	4.20×10^{-10}	4.20×10^{-10}	5.86×10^{-10}
Under	Expected Failure Rate (h^{-1})	2.57×10^{-5}	3.86×10^{-5}	4.40×10^{-5}	5.14×10^{-5}
	% Error	-43.8%	-15.6%	-3.7%	+12.5%
	"True" Value in 80% Credible Interval?	No	Yes	Yes	Yes
	Variance in Failure Rate (h^{-2})	1.32×10^{-10}	3.54×10^{-10}	4.79×10^{-10}	5.86×10^{-10}

accumulated either by operating one unit at a time for a very long time, or by operating multiple units in parallel for a long time.

Table 1 summarizes the results at the end of the 10-year simulated test, including the final expected failure rate estimate, the percent error of that estimate relative to the "true" value used for test simulation, whether or not the "true" value is contained within the 80% credible interval, and the variance in the updated estimate. If the prior failure rate estimate is accurate, Bayesian updating from that prior generates very good results. However, it is unlikely that the prior failure rate is accurate. Analysis of changes in failure rate estimates for ISS systems has shown that prior failure rates were often significantly different from current estimates. Across all ISS ORUs that have experienced a Bayesian update, approximately 85% of items had failure rates that were initially overestimated (i.e. the ORU turned out to be more reliable than expected), while 15% of prior failure rates were initially underestimated (i.e. the ORU turned out to be less reliable than expected). These errors in initial priors could be significant, with the updated estimates being several times larger or smaller than the initial estimate.[4] The prior failure rate was overestimated for approximately half of ECLSS ORUs, and for a quarter of ECLSS ORUs the current failure rate estimate is less than half of the original estimate. On the other hand, 18% of ECLSS ORUs had a prior failure rate that was lower than the current estimate, and 12% have current failure rates that are more than twice as much as was initially estimated.[7]

When the prior failure rate is an overestimate, the equal error factor approach quickly corrects this overestimate and results in a much more accurate estimate after 10 years. In contrast, the V2E ratio approach is slower to correct the overestimate, and the variance approach is much slower to correct, with over 200% error in the estimate even after a decade of testing. Correcting overestimated failure rates is important since they would otherwise result in excess spares for those items, which would increase total spares mass and potentially shift the spares balance away from other ORUs that would have benefitted from additional spares. Effectively, the V2E ratio and variance approaches are conservative, since overestimated failure rates are more likely to remain overestimates.

When the prior failure rate is an underestimate, however, the equal error factor approach struggles to correct the error, while the V2E and variance approaches are more effective. Under the simulated data above, if the prior failure rate were underestimated by a factor of 5 (which historical data indicate is possible[4,7]) then the estimate would still be underestimated by nearly 44% after 10 years of testing, and the true failure rate would not lie within the 80% credible interval of the updated failure rate estimate. Put another way, the ORU may actually be significantly less reliable than is indicated by a Bayesian update from test data when equal error factors are applied. The mission impact of this underestimated failure rate is that the POS associated with the spares allocation for that ORU will be less than expected, and mission risk will be higher. This mirrors results from previous work.[6]

The Jeffreys prior does not use prior failure rate estimates, and as a result is not impacted by over- or underestimated initial failure rates. For the simulated test data above, the Jeffreys prior generates a posterior failure rate estimate with less than 13% error relative to the true value. As noted in Section II.C, this is considered a conservative estimation technique, and these results show that it tends to overestimate failure rate. However, the amount of overestimation decreases with test time.

In all cases, Bayesian updating is effective at reducing the level of variance in the estimate. For example, the initial variance in the accurate prior case was 2.16×10^{-9}h^{-2}. After updating, the variance was reduced to less than 20% of this initial value. The variance after updating using the Jeffreys prior is approximately 40% higher than the results from cases that used the prior, at 5.86×10^{-10}h^{-2}, but this is still significantly lower than the initial variance (though of course the Jeffreys prior does not make any use of that initial estimate, so this is not a direct comparison). The percent reductions for over- and underestimated cases all depend on the initial variance, which is a function of the initialization method used, making direct comparisons more complicated. However, in all cases the variance after updating is on the same order of magnitude as the results from the accurate prior case and the variance using the Jeffreys prior.

It is important to note, however, that low variance is not necessarily correlated with accurate estimates. Accuracy and precision are two separate metrics, and while some prior initialization techniques produce updated estimates with low variance, they can also be highly inaccurate estimates. For example, when the equal error factor method is used with an underestimated prior, the resulting expected failure rate estimate is just over half of the true value. However, the variance is the lowest of all of the cases examined here, at 1.32×10^{-10}h^{-2}. This indicates a high risk for failure rate estimation that blindly follows one method: not only is the updated estimate incorrect by a significant amount, but it also indicates a high level of confidence in the incorrect estimate.

B. Sensitivity to Inaccurate Prior

Another way to examine the different prior initialization techniques is to calculate the expected updated (i.e. posterior) failure rate as a function of the given prior and an assumed "true" failure rate. Specifically, given a "true" failure rate λ_{true}, prior parameters α_{pri} and β_{pri}, and test time t_o, the expected failure rate after Bayesian updating is

$$E[\bar{\lambda}_{post}] = \frac{\alpha_{pri} + \lambda_{true} t_o}{\beta_{pri} + t_o} \tag{13}$$

since the expected number of failures during the test can be calculated based on the "true" failure rate and the amount of test time. In practice, this true failure rate cannot be known ahead of time, and the expected number of failures will often not be a whole number and therefore will not represent a valid test outcome. However, this approach enables the examination of the expected value of testing under an assumed "true" failure rate, which can help guide test planning and set expectations for test outcomes.

The difference in expected updated failure rate as a function of prior initialization methodology is driven by the way each method calculates α_{pri} and β_{pri} given an initial expected failure rate $\bar{\lambda}_{pri}$ and a parameter for setting initial uncertainty levels – either error factor ε_{pri}, V2E ratio γ, or variance σ^2_{pri}. When the Jeffreys prior is used, α_{pri} and β_{pri} are set directly. The equations for expected updated failure rate for each methodology are summarized in Table 2, and the expected updated failure rate for a range of values surrounding the assumed "true" value used for the simulated test results above. Figure 3 shows the expected updated failure rate for each case across a range of prior failure rates. Black horizontal and vertical dotted lines indicate the "true" failure rate, while gray vertical dotted lines indicate the over- and underestimate cases from above. In this case, since the expected number of failures after 10 years is a whole number, and the simulated test results resulted in that number, the expected updated failure rate shown here exactly matches the results from simulation discussed in Section A.

Table 2: Equations for the expected updated (posterior) failure rate estimate as a function of an assumed true failure rate λ_{true} and test time t_o.

Method	Expected Updated Failure Rate
Equal Error Factor	$E[\bar{\lambda}_{post}] = \dfrac{\left(e^{\left(\frac{\ln \varepsilon_{pri}}{1.645}\right)^2} - 1\right)^{-1} + \lambda_{true} t_o}{\bar{\lambda}_{pri}^{-1}\left(e^{\left(\frac{\ln \varepsilon_{pri}}{1.645}\right)^2} - 1\right)^{-1} + t_o}$
Equal V2E Ratio	$E[\bar{\lambda}_{post}] = \dfrac{\frac{\bar{\lambda}_{pri}}{\gamma_{pri}} + \lambda_{true} t_o}{\frac{1}{\gamma_{pri}} + t_o}$
Equal Variance	$E[\bar{\lambda}_{post}] = \dfrac{\frac{\bar{\lambda}^2_{pri}}{\sigma^2_{pri}} + \lambda_{true} t_o}{\frac{\bar{\lambda}_{pri}}{\sigma^2_{pri}} + t_o}$
Jeffreys Prior	$E[\bar{\lambda}_{post}] = \dfrac{0.5 + \lambda_{true} t_o}{t_o}$

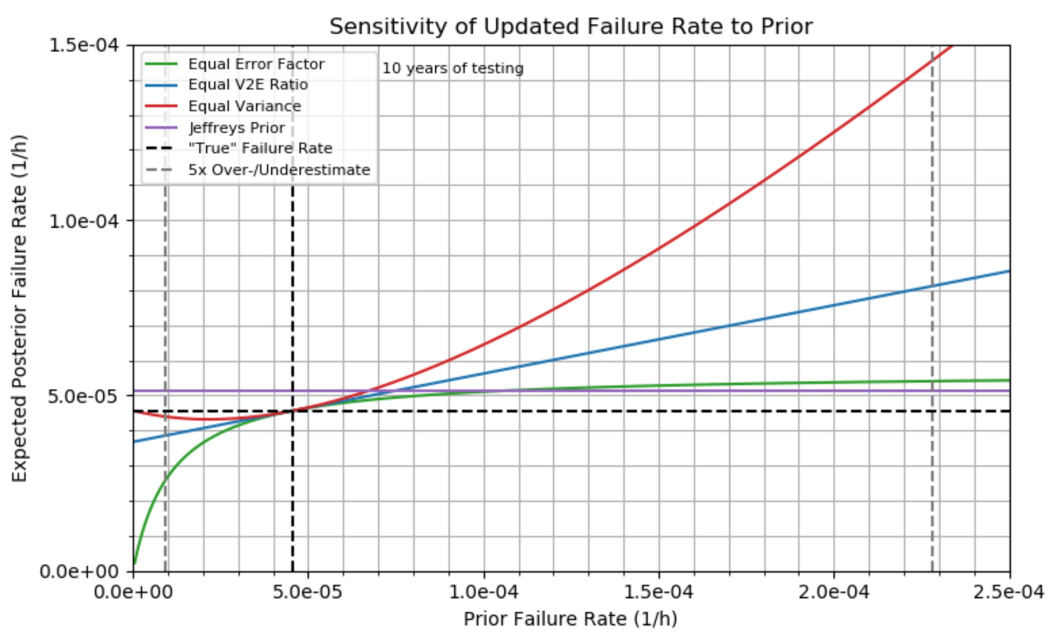

Figure 3: Sensitivity of expected updated (posterior) failure rate estimate to changes in the prior estimate, using the notional failure rate and 10-year test described in Section III.A.

Figure 3 shows that the different prior uncertainty initialization methods produce very different responses as a function of the error in the prior estimate. The equal error factor method essentially initializes the prior with variance proportional to the square of the initial failure rate estimate. As a result, it assigns significantly more variance to higher failure rate estimates and is more effective at correcting them downwards; however, it also assigns significantly less variance to lower failure rates estimates, and is less effective at correcting them upwards. These trends are illustrated by the green curve, which shows error increasing slowly when the prior failure rate estimate increases from the true value (indicated by the black dotted lines), but increasing quickly when the prior decreases from the true value. The equal variance method (red) shows the opposite behavior. Error grows much faster when failure rates are overestimated, but underestimated failure rates are quickly corrected and error is lower. The equal V2E ratio method, in contrast, produces an even response across the entire spectrum, showing a linear response in both directions, with the slope determined by both the V2E ratio γ and the test time t_o. That is, the V2E method is equally effective for correcting overestimates and underestimates. Jeffreys prior, as noted earlier does not use the prior estimate and therefore is not sensitive to the accuracy of prior estimates.

IV. Impact of Initialization Method on Spares Mass Estimates

Section III highlighted the impact of different prior initialization methods on the updated failure rate estimate for a notional ORU. In order to explore the impact that these different methods can have on spares mass, this section examines how spares mass requirements for a notional 1,100-day Mars mission are impacted by different estimating methods. A subset of ISS ECLSS is modeled, consisting of the Water Processor Assembly, Urine Processor Assembly, Oxygen Generation Assembly, Carbon Dioxide Removal Assembly, Common Cabin Air Assembly, and Trace Contaminant Control System, and spares allocations and resulting spares mass are calculated assuming a required POS of 0.98. These systems were

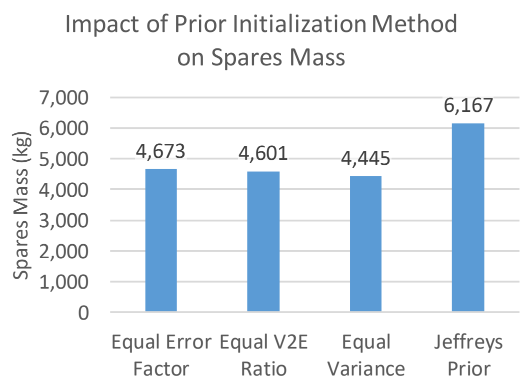

Figure 4: Total spares mass for a subset of ECLSS for a 1,100 day mission with a POS of 0.98 as a function of prior uncertainty initialization method.

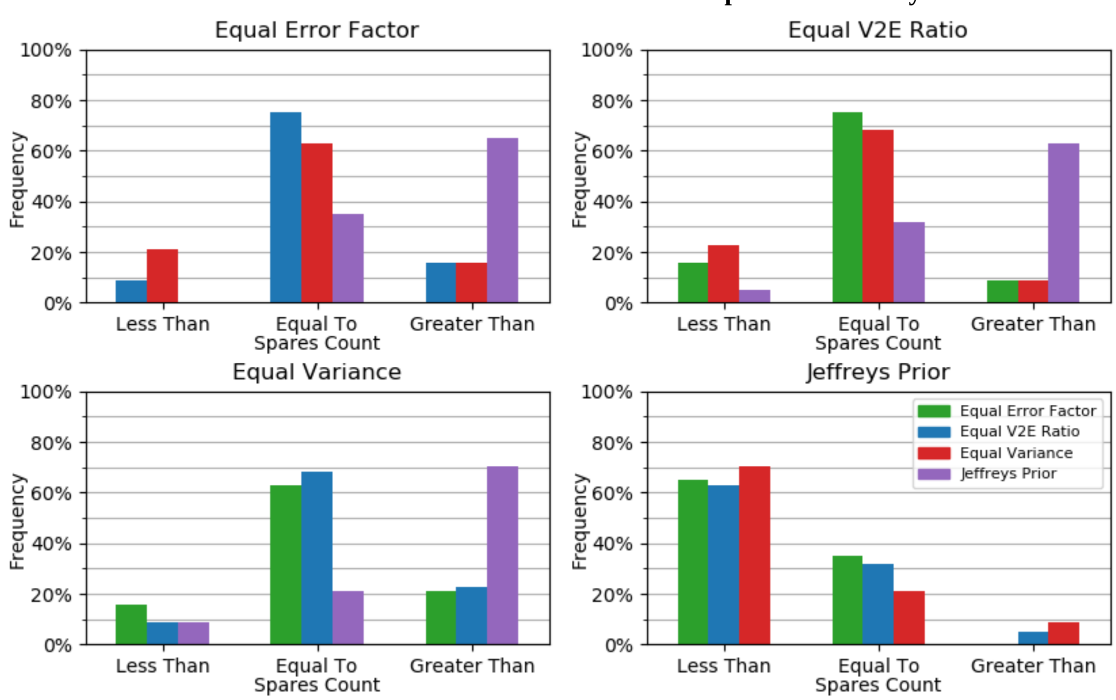

Figure 5: Histograms showing the percentage of spares for which the spares count under a particular model (indicated by color) is less than, equal to, or greater than the spares count under another model (indicated by axis title).

selected for this analysis due to ready availability of operating history data for those systems, including accumulated operating hours and number of random failures for each ORU. Failure rate estimates are generated for each ORU based on that ORU's initial failure rate estimate, operating history, and the method used to initialize the level of uncertainty in the prior. Note that this notional case study is for illustrative purposes, and is not meant to represent actual spares loads for a Mars mission.

Figure 4 shows the total spares mass for the systems modeled here as a function of prior initialization technique. The spares mass resulting from use of the Jeffreys prior for all ORUs is higher than all options, as expected. The Jeffreys prior is a conservative failure rate estimation technique, and therefore is likely to slightly overestimate all failure rates, as indicated in the analysis described in Section III above. As a result, more spares are allocated to achieve the POS target, and the total spares mass is higher. The other estimating techniques fall within a few hundred kilograms of each other. The equal error factor approach producing the next highest total spares mass, followed by the equal V2E ratio approach. The equal variance approach results in the lowest total spares mass.

It is important to note, however, that allocating sufficient spares mass is not the same as allocating sufficient spares. The value that matters for POS and risk assessment is how many spares are provided for each ORU individually. Figure 5 shows, for each model, the percentage of spares for which the spares count was less than, equal to, or more than the spares count for that same spare under a different model. For example, the blue bars in the upper left quadrant of Figure 5 indicate the less than / greater than / equal to distribution for spares allocations based on the equal V2E method, relative to the equal error factor method. These results show that even for the two cases that resulted in the most similar spares allocations (i.e. equal error factor and equal V2E ratio), only 75% of ORUs had the same number of spares provided to them. For 16% of ORUs, the number of spares provided under the equal error factor model was less than the number provided under the equal V2E ratio model. If, for example, the failure rate estimates resulting from applying the equal V2E ratio were more accurate than those resulting from applying the equal error factor method, then the POS associated with those ORUs would not be as high as expected, and the overall mission would likely have more risk. This example is for illustrative purposes only, and is not meant to say that the V2E ratio approach necessarily produces more accurate failure rate estimates than the equal error factor approach. However, it serves as an example of how differences in failure rate estimates resulting from different prior initializations can significantly impact mission risk, even if the overall spares mass estimates are similar.

V. Discussion

The notional case studies presented in Sections III and IV show that failure rate estimates are strongly impacted by the choice of prior initialization methodology, even after years of testing. Since initial failure rate estimates are often not accurate, it is critical that the selected methodology be effective at updating inaccurate priors. The equal error factor approach assigns variance to prior estimates proportional to the square of the initial failure rate estimate, and as a result is very effective for updating overestimated priors, but is extremely slow to correct underestimates. In contrast, the equal variance method quickly corrects underestimates, but is extremely slow to correct overestimates. The equal V2E ratio method assigns variance proportional to the initial failure rate estimate, and is equally effective for updating overestimates or underestimates, providing a linear response across the range of potential priors. The Jeffreys noninformative prior does not use prior estimates and is therefore not impacted by the accuracy of those estimates, but tends to result in a conservative (i.e. overestimated) failure rate estimate.

The true failure rate of each ORU will never be known with certainty. All of these methods are simply techniques for generating estimates based on an initial guess and test data. The resulting estimates will never be perfect, but they will drive risk and spares mass for future missions, and therefore it is important that these estimates be as accurate and precise as can be achieved under reasonable constraints. In addition, any potential bias in the estimating technique must be considered when using results from that technique. Given an initial failure rate estimate population that contains both overestimates and underestimates, some approaches are more likely to correct overestimates than underestimates, and vice versa. The consequences of overestimates are different from those of underestimates, however. Overestimated failure rates are conservative – a belief that the ORU in question is less reliable than it actually is. This leads to an excess spares allocation for that item, but also a true mission risk that is less than the estimate. On the other hand, underestimated failure rates are overly optimistic – a belief that an ORU is more reliable than it actually is. While this results in a lower spares allocation for that item, it also means that true mission risk is greater than the estimate, sometimes significantly so. Programs should balance the mass impacts of potentially carrying too many spares as a result of conservative failure rate estimates against the cost of hidden risk that could result if failure rate estimates are too optimistic.

Under all methods, however, these results show that Bayesian updating can be a very slow process, especially when only one instance of the system is operated at a time. The simulated test results shown in Figure 1 and Table 1 show that even after a decade of testing the updated failure rate estimate may still be significantly inaccurate and uncertain. This process can be accelerated by building multiple copies of systems, but current ECLSS development practices tend to include only one or a handful of flight-like units for testing. In addition, test schedules are typically significantly shorter than the durations examined here. While this approach is workable for missions in LEO that have access to abort and resupply to mitigate risk, it is not an effective way to manage risk for long-endurance missions beyond LEO. Programs should carefully weigh the cost of purchasing additional copies of key systems for testing during development against the potential benefits that data gathered from those units could provide – not just in terms of reducing spares mass by reducing uncertainty, but also in terms of risk reduction through empirical verification of failure rate estimates. In addition, the operational experience gained by operating these units (which is gained faster as more units are operated in parallel) can help uncover potential failure modes within the system and provide the opportunity to remove them, improving the reliability of the system overall.

VI. Conclusions

Reliability estimation is difficult. Prior estimates are often inaccurate, and though Bayesian techniques can be used to update estimates based on observed data, these data can take a very long time to generate. In addition, updated estimates are strongly influenced by the method used to generate priors. If design life testing is conducted by operating only one or a handful of units in parallel – as is currently typically the case – many years of testing will be required to produce enough data to reduce uncertainty and verify that the updated estimate is close to the true value. However, testing multiple units in parallel provides a multiplier on the rate of data gathering that can significantly improve resulting failure rate estimates and reduce the dependence of the estimate on the prior. In addition, careful evaluation of the method used to generate priors can provide insight into the likely bias in the updated estimates and its associated impact on spares mass and risk estimates. Overall, the prior uncertainty initialization method, number of test units, and available test schedule are key drivers of a program's ability to drive uncertainty out of the system and verify failure rate estimates. These estimates will be stronger drivers of mass and risk for missions beyond LEO in the future, and therefore greater emphasis should be placed on reliability test and evaluation programs than has been in the past.

Acknowledgements

The authors would like to thank Chris Jones for his thoughts and input on this work.

References

[1] Owens, A., De Weck, O., Stromgren, C., Goodliff, K. E., and Cirillo, W. Supportability Challenges, Metrics, and Key Decisions for Future Human Spaceflight. 2017. AIAA SPACE 2017, Orlando, FL. AIAA 2017-5124.

[2] Cirillo, W., Aaseng, G., Goodliff, K., Stromgren, C., and Maxwell, A. Supportability for Beyond Low Earth Orbit Missions. Presented at the AIAA SPACE 2011 Conference & Exposition, Long Beach, CA, 2011. AIAA 2011-7231.

[3] Owens, A. C. *Multiobjective Optimization of Crewed Spacecraft Supportability Strategies*. Doctoral Thesis. Massachusetts Institute of Technology, Cambridge, MA, 2019.

[4] Stromgren, C., Goodliff, K. E., Cirillo, W., and Owens, A. The Threat of Uncertainty - Why Using Traditional Approaches for Evaluating Spacecraft Reliability Are Insufficient for Future Human Mars Missions. Presented at the AIAA SPACE 2016, Long Beach, CA, 2016. AIAA 2017-5307/

[5] Owens, A. C., de Weck, O. L., Stromgren, C., Goodliff, K., and Cirillo, W. Accounting for Epistemic Uncertainty in Mission Supportability Assessment: A Necessary Step in Understanding Risk and Logistics Requirements. Presented at the 47th International Conference on Environmental Systems, Charleston, SC, 2017. ICES-2017-109.

[6] Owens, A., and de Weck, O. How Much Testing Is Needed to Manage Supportability Risks for Beyond-LEO Missions? Presented at the International Conference on Environmental Systems, Boston, 2019. ICES-2019-66.

[7] Owens, A., and de Weck, O. International Space Station Operational Experience and Its Impacts on Future Mission Supportability. Presented at the International Conference on Environmental Systems, Albuquerque, NM, 2018. ICES-2018-198.

[8] Dezfuli, H., Kelly, D., Smith, C., Vedros, K., and Galyean, W. *Bayesian Inference for NASA Probabilistic Risk and Reliability Analysis*. Publication NASA/SP-2009-569. National Aeronautics and Space Administration, Washington, DC, 2009.

[9] Peter, A., Das, D., and Pecht, M. Appendix D: Critique of MIL-HDBK-217. In *Reliability Growth: Enhancing Defense System Reliability*, National Academies Press, pp. 203–245.

[10] Ebeling, C. E. *An Introduction to Reliability and Maintainability Engineering*. Waveland Press, Long Grove, Ill, 2010.

[11] Shynk, J. J. *Probability, Random Variables, and Random Processes: Theory and Signal Processing Applications*. Wiley, Hoboken, NJ, 2013.

[12] Karlis, D., and Xekalaki, E. "Mixed Poisson Distributions." *International Statistical Review*, Vol. 73, No. 1, 2007, pp. 35–58. doi:10.1111/j.1751-5823.2005.tb00250.x.

[13] Slay, F. M., Bachman, T. C., Kline, R. C., O'Malley, T. J., Eichorn, F. L., and King, R. M. *Optimizing Spares Support: The Aircraft Sustainability Model*. Publication AF501MR1. Logistics Management Institute, McLean, VA, 1996.

[14] Fox, B. "Discrete Optimization Via Marginal Analysis." *Management Science*, Vol. 13, No. 3, 1966, pp. 210–216. doi:10.1287/mnsc.13.3.210.
[15] Hillier, F. S., and Lieberman, G. J. *Introduction to Operations Research*. McGraw-Hill, Boston, 2001.