IAC-09-D1.3.6

STATISTICAL RELIABILITY ANALYSIS OF SATELLITES BY MASS CATEGORY: DOES SPACECRAFT SIZE MATTER?

Gregory F. Dubos
Georgia Institute of Technology, United States
greg.dubos@gatech.edu

Jean-Francois Castet
Georgia Institute of Technology, United States
jcastet3@gatech.edu

Joseph H. Saleh
Georgia Institute of Technology, United States
jsaleh@gatech.edu

ABSTRACT

Reliability has long been recognized as a critical attribute for space systems, and potential causes of on-orbit failures are carefully sought for identification and elimination through various types of testing prior to launch. From a statistical or actuarial perspective, several parameters of the spacecraft, such as mission type, orbit, or spacecraft complexity, can potentially affect the probability of failure of satellites. In this paper, we explore the correlation between satellite mass, considered here as a proxy for size, and satellite reliability, and we investigate whether different classes of satellite, defined in terms of mass, exhibit different reliability profiles. To this end, we first conduct nonparametric analysis of satellite reliability based on a sample of 1,444 satellites. The satellites are organized in three main categories defined by satellite mass (Small – Medium – Large). Three nonparametric reliability curves are thus derived. We then provide parametric fits of the reliability curves to facilitate the identification of failure trends. We proceed to the comparative analysis of failure profiles over time and clearly identify different reliability behaviors for the various satellite mass categories. Finally, we discuss possible structural and causal reasons for these trends and failure differences, in particular with respect to design, testing and procurement.

1. INTRODUCTION

Reliability has long been recognized as a critical attribute for space systems, and potential causes of on-orbit failures are carefully sought for identification and elimination through various types of testing prior to launch. Unfortunately, despite the recognition of its importance, limited on-orbit failure data and statistical analyses of satellite reliability exist in the technical literature. To help fill this gap, Castet and Saleh [1] recently collected failure data for 1,584 Earth-orbiting satellites launched between January 1990 and October 2008, and conducted statistical reliability analysis for this extensive sample. Nonparametric reliability results from their study along with the 95% confidence intervals are shown in Fig. 1.

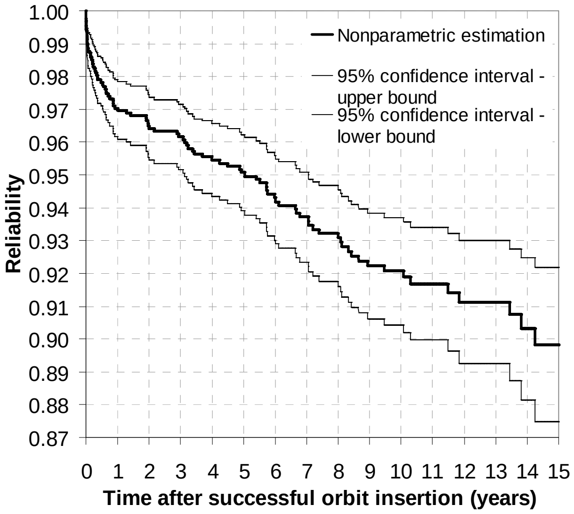


Fig 1. Satellite reliability with 95% confidence intervals (details in [1])

One limitation of the work in [1] is that satellites of different types and in different orbits have been lumped together, and their "collective" failure behavior was statistically analyzed. The issue with this approach is that no two (or more) satellites are truly alike, and, unless they are co-located, satellites operate in and experience different environmental conditions. As a result, the assumption that the failure times of the satellites are independent and identically distributed (*iid*) may be challenged, and the "collective" reliability results may not accurately reflect the specific reliability of a particular spacecraft or spacecraft type. The statistical analysis dilemma results from the fact that the space industry lacks "satellite mass production," and does not have the luxury of say the semi-conductor industry where data on thousands of identical transistors operating under identical environmental conditions can be available for statistical analysis, or other industries with products for which failure data can be easily obtained from accelerated testing. Given the relatively small number of satellite launched, data specialization for example for specific spacecraft platform may result in significantly reduced sample size thus constraining the statistical reliability analysis and its precision. In this work, we provide a first-order data specialization, by satellite mass category, and for which the sample size (and failure occurrence within the sample) remains appropriate for statistical analysis.

From a statistical or actuarial perspective, several parameters or characteristics of the spacecraft, such as the spacecraft complexity, its number of instruments or its payload size, to name a few, can potentially affect the probability of failure of satellites. In this paper, we explore the correlation between satellite mass and satellite reliability, and we investigate whether different classes of satellite, defined in terms of mass, exhibit different reliability profiles. We address the following questions: for example, are different spacecraft masses correlated with different failure behaviors on-orbit? Do small satellites exhibit different failure behaviors on-orbit, hence different reliability profiles, than larger one? And more broadly, do different satellite classes (in terms of mass) have different reliability profiles?

This possible correlation between spacecraft mass and reliability has not to date been investigated from a statistical perspective. Intuitive trends have often been discussed, but sometimes yielded contradictory conclusions. On one hand, an increase in mass has naturally been associated with the use of design redundancy on-board a spacecraft to improve reliability [2]. On the other hand, an increase in complexity (for which mass is often considered a good proxy) has long been seen as a factor degrading reliability. This phenomenon was already identified during the Apollo program: pressure-fed and storable propellants on the lunar module propulsion systems allowed bypassing the use of ignition systems and pumps, resulting in a reduction of mass and complexity, and ultimately an increase in reliability [3]. Furthermore, recent work by Bearden [4] tended to show that NASA spacecraft that failed were characterized by a high complexity factor (defined by the author as an average of technical factors including spacecraft mass). Finally, in [5], Fleeter discusses a simple model of spacecraft reliability $R = R_0^n$, where n is the number of components that "nominally scales with mass", and R_0 is the reliability of each component. Using this relation, the author argues that "using the same part quality, [a] little spacecraft will be more reliable" than a larger one, as n will be smaller. As noted by Sarsfield [6], the question of system size in relation to reliability divides the spacecraft community with on one side the proponents of small, "single-string", and thus simple systems, and on the other, the advocates of larger systems using more redundancy.

Quantitative answers have therefore to be found to resolve this issue and identify correlation, if any, between spacecraft mass and reliability. To this end, we conduct in this work statistical reliability analysis of satellites arranged by mass categories and we investigate whether these different classes of satellite exhibit different reliability profiles.

The remainder of this paper is organized as follows. In section 2, we present the data used in this work and introduce our classification of satellites based on mass. In section 3, we conduct a nonparametric analysis of satellite reliability for each class, using the Kaplan-Meier estimator (given the censored nature of the data). To facilitate the identification of trends in the reliability behavior, we then provide in section 4 parametric fits of the reliability curves using the Maximum Likelihood Estimation method as well as mixtures of Weibull distributions. In section 5, we proceed to the comparative analysis of failure profiles over time, which clearly identifies different reliability behaviors for different satellite mass bins. Finally, beyond the statistical identifications of these differences, we discuss in section 6 possible structural and causal reasons for these trends and failure differences, in particular with respect to design, testing and procurement practices.

2. DATABASE, DATA DESCRPITION AND CATEGORIZATION

For the purpose of this study, we used an extensive database of failures and anomalies on-orbit. Details about the database can be found in [1]. While not "complete" in a statistical sense, this database is widely used and considered the most authoritative in the space industry with failure data for over 6,400 spacecraft.

We restricted our study to Earth-orbiting satellites successfully launched between January 1990 and October 2008. As a result, we retained from the database 1,444 satellites launched within this time period and for which the satellite mass (at launch) was available. We used for our reliability calculations what is referred to in the database as a Class I failure, that is, a retirement of a satellite due to failure. For each spacecraft in our sample, we collect: 1) its mass; 2) its launch date; 3) its failure date, if failure occurred; and 4) the “censored time”, if no failure occurred. This last point is further explained in the following subsection 3.1. The data collection template and sample data for our analysis are shown in Table 1.

Table 1. Data collection template and sample data for our statistical analysis of satellite reliability

Sample unit number	**Mass at launch (kg)**	**Launch date**	**Failure date (if failure occurred)**	**Censored time (if no failure occurred)**
Satellite #1	1500	11/06/1998	11/15/1998	–
Satellite #2	480	03/01/2002	–	10/02/2008
…		…	…	…
Satellite #1,444	2600	04/26/2004	03/28/2006	–

After the data was collected, we categorized the satellites into different mass bins. Various taxonomies for spacecraft based on their mass have been used over time in the space industry. For example, Sarsfield [6] points out that “there is no official definition of a small satellite”, even though this qualifier has gained much popularity during the last two decades. The author adds, “the Center for Satellite Engineering Research at the University of Surrey defines a “mini” satellite as being between 100 and 500kg.” Similarly, a list of satellites launched from 1991 to 1995, whose mass is under 425 kg and considered as “small”, is provided in [5]. The National Research Council’s Aeronautics and Space Engineering Board established a Panel on Small Spacecraft Technology that defined small spacecraft as those “weighing approximately 600 kg or less” [7]. In this work, we adopt the more commonly used definition of “small spacecraft” as those within the 0-500kg range [6]. This range corresponds to the categories AW and BW in the ANSI/AIAA guidelines for spacecraft design [8]. Furthermore, in these guidelines, the 500-2,500 kg bin corresponds to the category CW, and masses above 2,500 kg fall into the last category DW.

Based on the previous discussion, we retained for our analysis the classification of spacecraft that is presented in Table 2.

Table 2. Categories of spacecraft based on mass

Mass at launch (kg)	**Spacecraft category**	**Examples**
[0-500]	Small (S)	FAST, JASON 1, NANOSAT 01
]500-2500]	Medium (M)	TOPEX-POSEIDON, GPS NAVSTAR II-06
>2500	Large (L)	DirecTV 1R, HotBird 8

Note that the IRIDIUM series of satellites which were initially included the Medium category has been removed from our sample for the purpose of this statistical analysis. These satellites, based on the same design with a mass of 657 kg, experienced a very large number of failures, mostly attributed to a malfunction of the Attitude and Orbit Control System. This recurrent cause of failure challenged the assumption of independence of the failures. When analyzed separately, the IRIDIUM series exhibited a very different reliability behavior from the other satellites belonging to the same mass category. This design of satellites therefore represented an “outlier” that introduced a significant bias in the Medium category that could potentially result in flawed interpretations for the satellites of this size in general. For these reasons, the IRIDIUM series was removed from this category. However, the reliability behavior of the IRIDIUM satellites in relation to the unique design, manufacturing and testing practices characterizing this series of spacecraft remains a very interesting research direction that should deserve special attention in future work.

In total, we have 415 satellites in our Small mass bin, 554 in the Medium mass bin (IRIDIUM excluded), and 475 in the Large mass bin.

3. NON-PARAMETRIC ANALYSIS OF SATELLITE RELIABILITY BY MASS CATEGORY

Censored Data Sample and Kaplan-Meier estimator

Right-censoring occurs in statistical life data analysis when some items under observation are removed from the sample before their failure occurs, or when the experiments or observation window ends and some items are still operational (again, their failure is not observed). Our sample has a combination of these two types of censoring, and it contains items with staggered entries. This is known in statistical analysis as Type IV censoring or random censoring, and it means the following: 1) the satellites in our sample are activated at different points in time (i.e., the satellites are launched at different calendar dates) but all these activation times in our sample are known, 2) failures dates and censoring are stochastic, and 3) censoring occurs either because a satellite is retired from the sample before a failure occurs or because the satellite is still operational at the end of our observation window (October 2008). Censoring requires careful attention: deriving a reliability function from censored life data is not trivial, and it is important that it is done properly if the results are to be meaningful and unbiased. In this work, we adopt the powerful Kaplan–Meier estimator [9], which is best suited for handling the type of censoring we have in our sample. The derivation of the Kaplan-Meier estimator formula can be found in [1]. The Kaplan-Meier estimator of the reliability function with censored data is given by Eq. (1):

$$\hat{R}(t) = \prod_{\substack{\text{all } i \text{ such} \\ \text{that } t_{(i)} \leq t}} \hat{p}_i = \prod_{\substack{\text{all } i \text{ such} \\ \text{that } t_{(i)} \leq t}} \frac{n_i - 1}{n_i} \tag{1}$$

where:

$t_{(i)}$: time to ith failure (arranged in ascending order)

$$\hat{p}_i = \frac{n_i - 1}{n_i} \tag{2}$$

n_i = number of operational units right before $t_{(i)}$

= n – [number of censored units right before $t_{(i)}$]

– [number of failed units right before $t_{(i)}$]

Should there be ties in the failure times, say m_i units failing at exactly $t_{(i)}$—this situation is referred to as a tie of multiplicity m—then Eq. 2 is replaced by:

$$\hat{p}_i = \frac{n_i - m_i}{n_i} \tag{3}$$

If a censoring time is exactly equal to a failure time, a convention is adopted that assumes censoring has occurred immediately after the failure (that is, at an infinitely small time interval after the failure).

Results

The data organized in mass bins is now treated with the Kaplan-Meier estimator (Eq. (1)), and we obtain the Kaplan-Meier plot of satellite reliability for each mass category of spacecraft, shown in Fig. 2.

Fig. 2 reads as follows. For example, after a successful launch, satellite reliability for the Medium size category (500–2500 kg) drops to approximately 97% after five years on-orbit. More precisely, we have:

$$\hat{R}(t) = 0.968 \qquad \text{for} \quad 4.548 \text{ years} \leq t < 5.719 \text{ years}$$

Both small (≤ 500 kg) and large (> 2500 kg) satellites exhibit a reliability of 95% after two years. Past 12 years, satellite reliability drops to 87% for large satellites, 91% for small satellites, and 94% for medium satellites. Some important failure trends and difference between the three satellite mass categories can already be seen in Fig. 2:

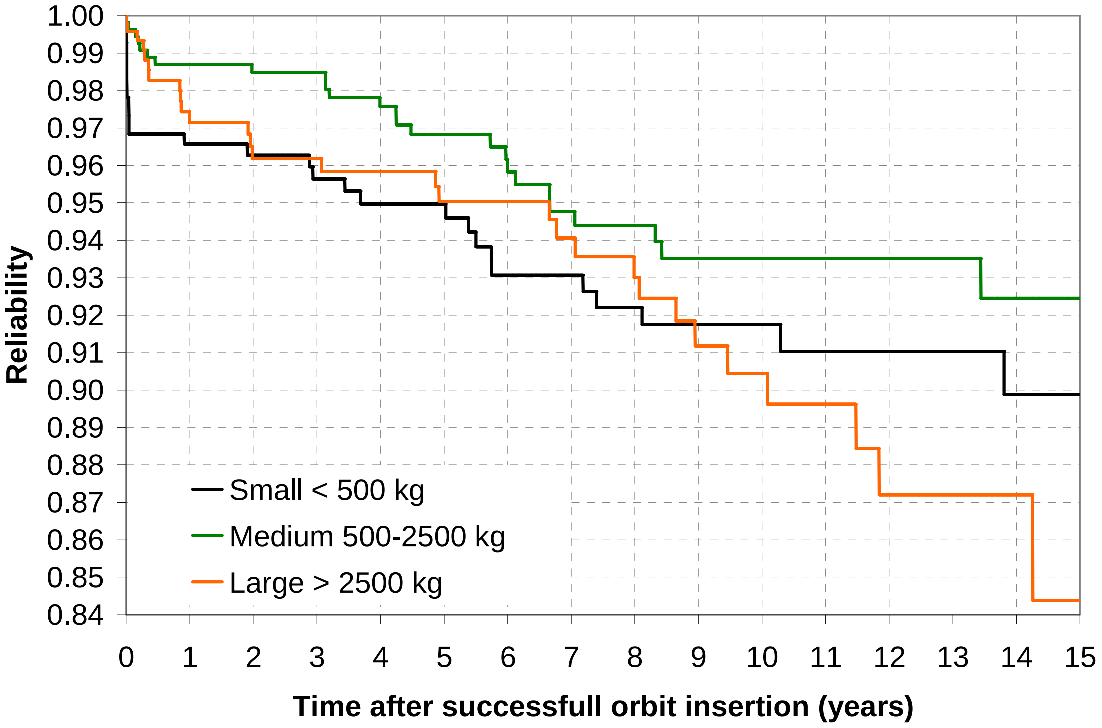


Fig. 2. Kaplan-Meier plot of satellite reliability for each mass category

- **Infant mortality**: small satellites (with a mass less than 500 kg) exhibit a significant drop in reliability during the first months following orbit insertion. For example, six months after orbit insertion, their reliability is already down to approximately 96.8%. This striking behavior of "infant mortality" experienced by small satellites is much less severe in the case of bigger satellites. Indeed, for the two other categories (Medium and Large), the drop in reliability is initially more moderate. Six months after the orbit insertion, the reliability of large satellites is around 98.3%, and that of medium satellites is 98.7%.

- **Wear-out**: while the drop in reliability of small satellites appears to taper off after 7 years, the large spacecraft (mass greater than 2,500 kg) exhibit a very different failure behavior after 7 years, marked by a steep decrease in reliability. For this Large category of satellites, this drop is more severe than during the first 6 years, as reflected by a change of convexity of the reliability curve approximately 6.5 years after orbit insertion. The "wear-out" failure behavior seems clearly more distinct for large satellites than smaller ones.

- Except for the difference in infant mortality, both the Medium and Small categories exhibit a similar reliability behavior, with a moderate decrease in the reliability from year 1 to 8 years, and a tapering off or shallower drop from year 8 to 15.

- In the sample of 1,444 satellites we analyzed, the Medium category exhibits the highest reliability of all satellites, always remaining above 92.4% over the course of 15 years after the orbit insertion.

- The Small category exhibits the lowest reliability of all satellites up to 9 years after orbit insertion. However, past 9 years, the large satellites reclaim the leadership in failure as their reliability steadily drops below that of the two other categories.

These trends will be revisited more formally and analytically in Section 5. The important result from Fig. 2 is that different satellite mass categories do indeed have different reliability profiles and failure behaviors. In addition, Figure 2 indicates that the question whether smaller or larger satellites are more (or less) reliable is ill-posed; it cannot be answered without the specification of a time horizon of interest (see previous note on leadership in failure).

4. PARAMETRIC ANALYSIS OF SATELLITE RELIABILITY BY MASS CATEGORY

Nonparametric analysis provides powerful results since the reliability calculation is not constrained to fit any particular pre-defined lifetime distribution. However, this flexibility makes nonparametric results neither easy nor convenient to use for different purposes, as often encountered in engineering design (e.g., reliability optimization). In addition, some trends and patterns are more clearly identified and recognizable with parametric analysis. In the following, we present two parametric methods based on the Weibull distribution to fit the nonparametric reliability of each mass category discussed previously.

Weibull distribution

The Weibull distribution is one of the most commonly used distribution in reliability analysis. Its reliability (or survivor) function can be written as follows:

$$R(t) = \exp\left[-\left(\frac{t}{\theta}\right)^{\beta}\right] \quad \text{for } t \geq 0 \qquad (4)$$

where β is the shape parameter (dimensionless) and θ the scale parameter (units of time), both nonnegative. The reason for the wide adoption of the Weibull distribution is that it is quite flexible, and with an appropriate choice of the shape parameter β, it can capture different kinds of failure behaviors. For example, when $0 < \beta < 1$, the Weibull distribution models infant mortality (which corresponds to a decreasing failure rate); when $\beta = 1$, the Weibull distribution becomes equivalent to the Exponential distribution (constant failure rate); and when $\beta > 1$, the Weibull distribution models wear-out failures (which corresponds to an increasing failure rate).

In previous publications, we demonstrated the appropriateness of the Weibull distribution as a parametric model for satellite reliability [1,10,11]. In this work, we first derive Weibull fits for the three nonparametric reliability results using the Maximum Likelihood (MLE) procedure. However, the parametric results will be shown to be within 1.8 to 3.5 percentage points of the "benchmark" nonparametric results, and for our purposes, these results are not sufficiently accurate. We therefore proceed with deriving mixture Weibull distributions for the nonparametric results and demonstrate a significant improvement in the accuracy of the parametric fits. The details are discussed next.

Maximum Likelihood Estimation (MLE) of single Weibull fit

Details of the Maximum Likelihood Estimation procedure can be found in [12], and its analytic derivation is provided in [10]. When applied to the nonparametric reliability results shown in Fig. 2, the MLE procedure yields the Weibull parameter estimates for each satellite mass category. The results are provided in Table 3.

Table 3. Maximum Likelihood Estimates of the Weibull parameters for each mass category of satellites

Mass category	β	θ
		years
Small (≤ 500 kg)	0.3224	21414.5
Medium (500 – 2500 kg)	0.5973	1469.2
Large (> 2500 kg)	0.6794	291.4

Consider for example the small satellite category. Given Equation 4 and the information provided in Table 3, its nonparametric reliability is best approximated by the following Weibull distribution:

$$R_{Small}(t) = \exp\left[-\left(\frac{t}{21414.5}\right)^{0.3224}\right] \quad (5)$$

The values of the shape parameter (β = 0.3224) and the scale parameter (θ = 21414.5) are the Maximum Likelihood Estimates.

With a shape parameter $\beta < 1$, the Weibull fits of satellite reliability provided in Table 3 capture the existence of infant mortality for each mass category of satellites. Notice that the value of the shape parameter increases monotonously as the satellite mass increases (i.e., 0.3224 < 0.5973 < 0.6794). This trend is in agreement with the comment made previously regarding the increased risk of infant mortality as satellite mass decreases that was observed on the nonparametric reliability curves.

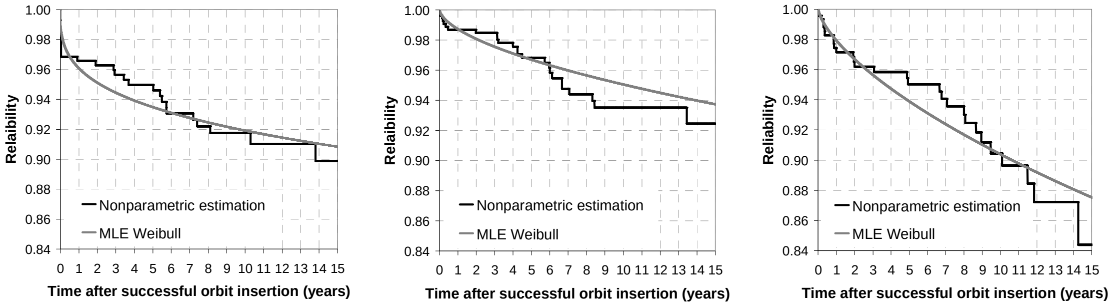
Fig. 3 shows the nonparametric reliability curve for the three mass categories, as well as the MLE Weibull fit. Fig. 3 provides a visual verification that the Weibull distribution with the MLE parameters provided in Table 3 is a good fit for the nonparametric reliability of large satellites.

For example, for the Large category, the maximum error (or distance) between the nonparametric reliability curve and the Weibull fit is 3.5 percentage points, and the average error is 1.1 percentage point. This represents a fair accuracy for a two-parameter (Weibull) distribution. Table 4 provides the maximum and average error between the nonparametric reliability and the Weibull fit for the three mass categories.

Table 4. Error between the nonparametric reliability and MLE Weibull fit for each satellite mass category

Mass category	Maximum error	Average error
	percentage point	percentage point
Small (≤ 500 kg)	1.8	0.7
Medium (500 – 2500 kg)	2.0	0.9
Large (> 2500 kg)	3.5	1.1

Table 4 shows that a single Weibull distribution provides a reasonable approximation of the nonparametric satellite reliability for each mass category, with an average error on the order of a single percentage point and a maximum error ranging from 1.8 to 3.5 percentage points. However, this reasonable approximation is not good enough for our purposes, and we can see for example in Fig. 3 that the parametric fit does not accurately follow the nonparametric ("benchmark") reliability results, especially between year 3 and 7, and between year 7 and 15 where clearly different failure trends are present. These different failure trends can be seen in the change of the convexity of the nonparametric curve around year 7, which reflects steeper failure propensity or reliability degradation after seven years on orbit. The single Weibull fit averages out these nuances and fails to capture these different failure trends. To improve the accuracy of the parametric fit, we derive next parametric fit with mixture distributions.



Small category Medium category Large category

Fig. 3. Nonparametric reliability and single Weibull fit for the three mass categories

Mixture distributions

Several distributions such as the Exponential, Weibull, or Lognormal, can be used as a basis for linear combination to generate a mixture distribution. In this subsection, we maintain the Weibull as the basis for our parametric calculations and derive mixture of two Weibull distributions for the nonparametric satellite reliability of each mass category. The parametric reliability model with a mixture of two Weibull distributions can be expressed as follows:

$$R(t) = \alpha \exp\left[-\left(\frac{t}{\theta_1}\right)^{\beta_1}\right] + (1-\alpha)\exp\left[-\left(\frac{t}{\theta_2}\right)^{\beta_2}\right] \qquad (6)$$

The parameter α is used to modify the relative weight given to each Weibull distribution in the mixture. A generalized expression for n mixture distributions is provided in [13]. We restrict our calculations in this work to $n = 2$ since as will be shown shortly, the results are significantly accurate and the 2-Weibull distributions follows with notable precision the different failure trends in the nonparametric results. Increasing n provides insignificant accuracy improvement.

The nonlinear least squares method provides us with the best fits for the parameters of the 2-Weibull mixture distribution for each mass category. The results are provided in Table 5.

For example, the resulting reliability function for large satellites is then expressed according to (Eq. 6), using the appropriate parameters of Table 5, as follows:

$$R(t) = 0.905 \exp\left[-\left(\frac{t}{24700}\right)^{0.3558}\right] + 0.095 \exp\left[-\left(\frac{t}{11.9}\right)^{3.579}\right] \qquad (7)$$

Notice that for the three satellite mass categories, the mixture distribution consists of a Weibull distribution capturing infant mortality ($\beta_1 < 1$), and another one capturing wear-out failures ($\beta_2 > 1$).

Table 5. Model parameters of the 2-Weibull mixture distribution for each mass category of satellites

Parameter	**Mass category**		
	Small (≤ 500 kg)	Medium (500 – 2500 kg)	Large (> 2500 kg)
α	0.9607	0.9703	0.905
β_1	0.2101	0.5071	0.3558
β_2	2.754	5.538	3.579
θ_1	10^7	6840	24700
θ_2	7.3	6.6	11.9

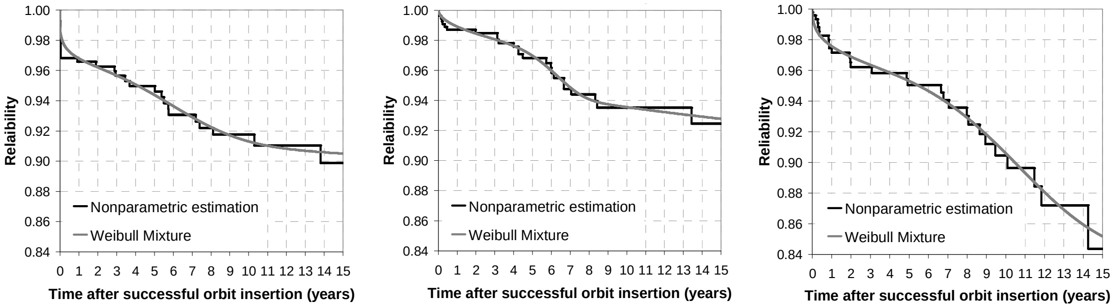
In addition, the infant mortality component of the mixture distribution has a significantly larger weight (α) than the wear-out component ($1 - \alpha$).

For the three mass categories, the new parametric fit of the reliability using a 2-Weibull mixture distribution accurately follows the nonparametric reliability, as shown in Fig. 4.

Table 6 provides the R^2 coefficients as well as the sum of the squares of errors (SSE) of the mixture distributions fits for the three satellite mass categories. In addition to the graphical inspection of the fits (like in Fig. 4), the high value of the R^2 (greater than 0.97) and the low value of the SSE in each case indicate that the fits obtained with the mixture distributions are significantly accurate.

Table 6. Measures of goodness-of-fit of the 2-Weibull mixture distribution
for each mass category of satellites

Coefficient	**Mass category**		
	Small (≤ 500 kg)	Medium (500 – 2500 kg)	Large (> 2500 kg)
R^2	0.9757	0.9841	0.9835
SSE	0.09438	0.06152	0.2017



Small category Medium category Large category

Fig. 4. Nonparametric reliability and 2-Weibull mixture fit for the three mass categories

To gauge the precision improvement between the single Weibull and the 2-Weibull mixture distributions, we calculate both the maximum and the average error between the nonparametric reliability (the benchmark results) and the parametric models. The results are shown in Table 7.

Table 7. Error between the nonparametric reliability and the parametric models over 15 years

Mass category	**Error**	**Parametric fit**	
	percentage point	Single Weibull	2-Weibull mixture
Small (≤ 500 kg)	maximum error	1.8	1.5
	average error	0.7	0.3
Medium (500 – 2500 kg)	maximum error	2.0	0.6
	average error	0.9	0.2
Large (> 2500 kg)	maximum error	3.5	1.5
	average error	1.1	0.4

As seen in Table 7, the 2–Weibull mixture distribution is significantly more accurate than the single Weibull distribution in capturing the (benchmark) nonparametric satellite reliability. For all mass categories, the average error for the 2–Weibull mixture distribution is reduced by over 50% compared with the average error of the single Weibull fit.

In the next section, we use these mixture distributions to further probe the difference between the failure behaviors and reliability trends of the three satellite mass categories.

5. COMPARATIVE ANALYSIS OF SATELLITE RELIABILITY ACROSS MASS CATEGORIES

In this section, we revisit the discussion in Subsection 3.2 regarding the difference in the reliability results of satellites in different mass categories. Fig. 5 shows the failure rates (or hazard function) of the Small and Large satellite mass categories. The y-axis is provided in log-scale for readability purpose. The upper panel in Fig. 5 provides a closer look at the failure rate over the short time periods (through the use of a log-scale on the x-axis).

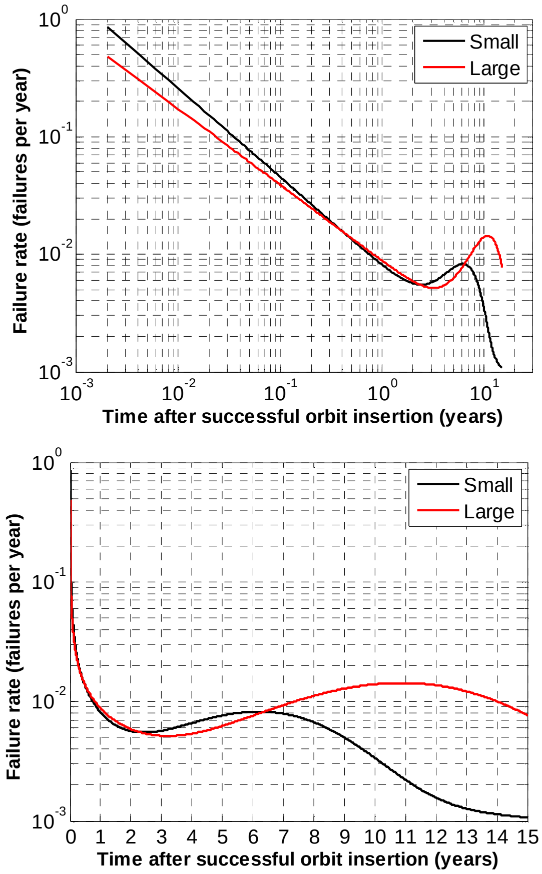


Fig. 5. Failure rates of Small and Large satellite mass categories

The failure rate $\lambda(t)$ uniquely determines the reliability function through Equation 8:

$$R(t) = e^{-\int_0^t \lambda(t')dt'} \qquad (8)$$

We first notice, on the upper panel in Fig. 5, that the failure rate of the small satellites is higher than that of the large satellite until roughly for the first four months on-orbit. This result reflects a previous observation following Fig. 2 that small satellites exhibit a more pronounced infant mortality than larger ones (small satellites exhibit a more significant drop in reliability over the first few months than larger ones, as shown in Fig. 2). We also observe, on the lower panel in Fig. 5, that the failure rate of the large satellites overtakes that of the small satellites around year 6.5. As a result, more distinct wear-out failures occur in large satellites than in small one (this is reflected in the change in convexity of the large satellites reliability in Fig. 2 around this same time). Given Equation 8 and the comparative shapes of the failure rates in Fig. 5, it is only a matter of time before the reliability of the large satellites drops below that of the small ones. This indeed can be seen to occur around year 9 in Fig. 2.

Fig. 6 shows the absolute difference (in percentage points) in satellite reliability for each pair of mass categories, namely Small/Medium, Small/Large and Medium/Large. This figure is rich in information, but should be interpreted with caution. For example, notice that the Small and Medium categories exhibit the largest difference in reliability of all the pairs up to nine years. This difference originates in the early life of the spacecraft, reflecting a difference in the infant mortality experienced by small satellites compared to medium satellites (as seen in Fig. 2 and Fig. 6). After one year, the curves indicate that the difference in reliability between those two categories remains stable, varying by less than 0.5 percentage point. In other words, the conditional reliabilities of small and medium satellites are roughly identical if they survive the first year on-orbit (the conditional reliabilities are explored in the next paragraph). By contrast, notice that while the reliabilities of the Small and Large categories remain somewhat "stable" within one percentage point for roughly the first 7 years, the two reliabilities diverge significantly after 9 nine years, suggesting a very different failure behavior between these two satellite categories during this time interval.

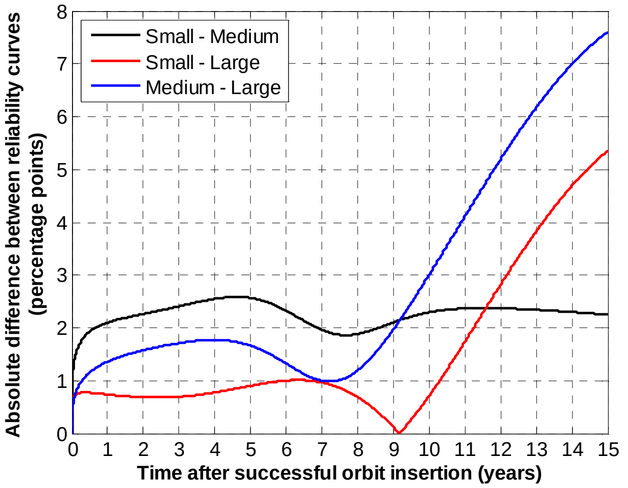


Fig. 6. Pairwise differences in satellite reliability over time

To better assess whether the reliability and failure behavior of two different mass categories are similar after a given period, we investigate their conditional reliabilities. For an item that has survived until time *T*, the conditional reliability allows the calculation of its probability of survival for an additional period of operation, knowing that the item has survived until *T*. By considering conditional reliabilities, we can perform a comparative analysis of failure behavior of the different satellite categories over different time periods and by selectively filtering out or disregarding failures prior to *T*. The benefits of doing so will be demonstrated shortly. Using the time domains shown in Fig. 7, the conditional reliability is defined as follows [14,15]:

$$R(t|T)=\Pr\{T_F>T+t|T_F>T\} \tag{9}$$

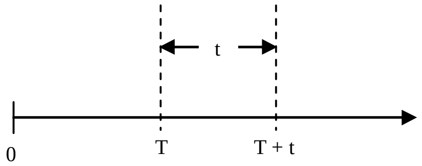


Fig. 7. Time domains for conditional reliability from [15]

T_F is the random variable Time-to-Failure. By definition of the conditional probability and the reliability function, (Eq. 9) can be reduced to:

$$R(t|T)=\frac{\Pr\{T_F>T+t\}}{\Pr\{T_F>T\}}=\frac{R(T+t)}{R(T)} \tag{10}$$

The conditional reliability is particularly useful for the study of a burn-in and its impact [16, 17]. In our case, we make a related, although broader, use of conditional reliabilities to study the failure behavior of satellites in different mass categories. The conditional reliability is useful for comparing two different reliability curves. Indeed, the conditional reliability "eliminates" or filters out the failure behavior of the system up to the time *T*. To illustrate the relevance of this observation for our study purposes, consider the following two systems, the first one suffering from significant infant mortality during the [0; t_1] period, and the second one is not. In addition, the two systems have the same failure behavior during the [t_1; t_2] period. The reliability curves of these two systems will be different and hardly comparable. While the reliability curves will clearly indicate the difference in infant mortality behavior between the two systems, these curves will not identify the similarity in failure behavior between the two systems during the [t_1; t_2] period. The difference between the curves is only due to the failures during the initial [0; t_1] period. Thus, by setting $T = t_1$, in Eq. 10, we can calculate the two conditional reliability curves over [t_1; t_2], and the two resulting curves will be similar, due to the same failure behavior during this period. By filtering out the failures during the initial period, the similarity of the failure behavior of the two systems during [t_1; t_2] can thus be clearly identified. Hence, by carefully selecting the appropriate time(s) *T*, the conditional reliability helps us separate the impact of early failures, and clearly determine periods of similar failure behavior, if they exist.

We performed an extensive scan of different values of *T* and retained the instants that yielded the most meaningful and relevant comparative analysis. For example, for the Small and Large satellite categories, Fig. 6 showed that their reliability behavior differed significantly at least until $t = 0.5$ year (sudden increase in the absolute difference of the reliability). Fig. 8a shows the absolute difference between the conditional reliabilities evaluated at $T = 0.5$ year for the Small and Large satellite categories. Fig. 8b clearly shows that after filtering out infant mortality or failures up to the first 6 months on orbit, the Small and Large categories have a similar failure behavior or conditional reliability profile until *t* = 8 years (with a small "bump' of less than 0.3 percentage points between year 4 and 8).

After *t* = 8 years, the absolute difference in conditional reliability increases suddenly, suggesting the divergence of the failure behaviors of the two satellite categories. Fig. 8b shows the actual conditional reliabilities evaluated at *T* = 0.5 year for the Small and Large categories. The two reliability curves overlap significantly between t = 0.5 and t = 8 years, confirming a similar failure behavior during this time period. At *t* = 8 years, the Large satellite category exhibits a much more severe decrease in reliability compared with that of the Small category.

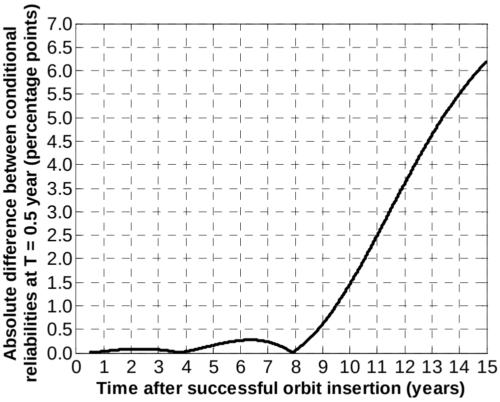


Fig. 8a. Absolute difference in conditional reliability evaluated at *T* = 0.5 year between the Small and Large categories

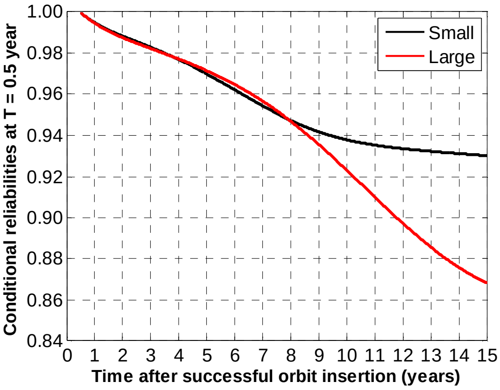


Fig. 8b. Conditional reliabilities evaluated at *T* = 0.5 year for the Small and Large categories

In summary, the statistical analysis revealed three periods of interest for the comparative reliability analysis of the Small and Large satellite categories. These periods and the failure behaviors in each period are provided in Table 8.

Table 8. Summary of reliability profiles for the Small and Large satellite categories

Category	**Period**		
	0 to 0.5 year	**0.5 to 8 years**	**8 to 15 years**
Small satellites	More pronounced infant mortality	Identical failure behavior	
Large satellites			Distinct wear-out failures

Fig. 9a shows the absolute difference between the conditional probabilities evaluated at *T* = 1.5 year for the Small and Medium satellite categories. The figure clearly shows that after filtering out infant mortality or failures up to the first years and a half on orbit, the Small and Medium satellite categories have a similar failure behavior or conditional reliability profile up to *t* = 15 years (with a small "bump' of less than 0.5 percentage points around year 5).

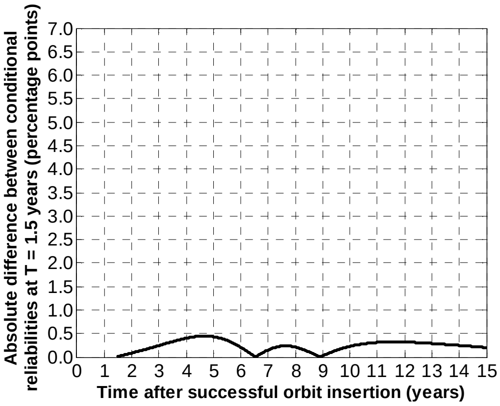


Fig. 9a. Absolute difference in conditional reliability evaluated at *T* = 1.5 years between the Small and Medium categories

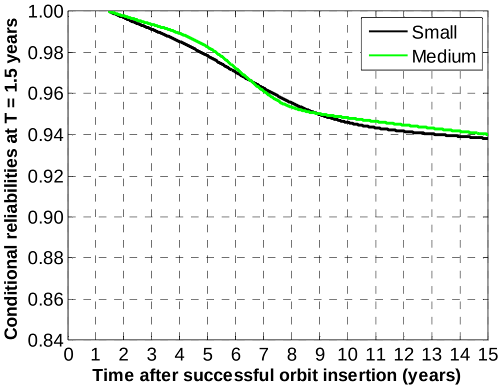


Fig. 9b. Conditional reliabilities evaluated at *T* = 1.5 years for the Small and Medium categories

Fig. 9b shows the actual conditional reliability curves. The two reliability curves overlap significantly between t = 1.5 and t = 15 years, confirming a similar failure behavior during this time period. Before T = 1.5 years, Fig. 6 had shown that the early failure behavior was relatively different, with a higher infant mortality for small satellites than for medium satellite.

The two periods of interest in the comparison of the reliability profiles for the Small and Medium categories are summarized in Table 9.

Table 9. Summary of reliability profiles for the Small and Medium satellite categories

Category	**Period**	
	0 to 1.5 year	**1.5 to 15 years**
Small satellites	More pronounced infant mortality	Identical failure behavior
Medium satellites		

Finally, visual inspection of Fig. 6 and analysis of the Medium – Large pair yield the following results: large satellites experience a more pronounced infant mortality than medium satellites up to 1.5 year, and they face more severe wear-out than medium satellites between 8 years and 15 years. Between 1.5 year and 8 years, the two satellite categories exhibit identical failure behavior.

In this section, we identified statistical difference in the failure behaviors of the three satellite categories. Our approach was based on observations of on-orbit failures and empirical calculations of on-orbit reliability. No causal analysis was attempted to ascertain the reasons why these differences in failure behaviors exist. Such an analysis is significantly wide-ranging and would require a dedicated monograph to be treated thoroughly. While such an effort is beyond the scope of this work, we do provide in the following section a set of hypotheses that may address the causal factors for the statistical differences in failure behavior of the three satellite categories.

6. HYPOTHESES FOR CAUSALITY ANALYSIS

Possible causes for the differences in failure behaviors identified in the previous section include factors related to the testing phase of the satellite prior to launch, to procurement and parts selections for the spacecraft, and to factors intrinsically related to the design type (size of the spacecraft)[*], as discussed in the following.

Testing

Small satellites do not benefit from the large budgets allocated to larger missions. Resource restrictions may limit the extent of the testing that is performed on small satellites. Indeed, the procedures, facilities and equipment (such as thermal-vacuum chambers) used to test a spacecraft often remain the same regardless of the size of the spacecraft [6]. In the case of small missions, extensive testing may thus have to be forfeited in order to meet budgetary constraints. Testing techniques such as parts “burn-in” are however critical to “remove latent defects and early failures” (if performed at appropriate stress levels and under proper environmental conditions [18]). The higher infant mortality exhibited by the smaller satellites may be due in part to differences in this final quality control gate that is testing, as a result of which potential early failures are screened, detected, and fixed for large satellites prior to launch. By contrast, smaller satellites, we hypothesize, exhibit more pronounced infant mortality because they may be subject to less stringent and extensive testing.

Procurement and parts selection

Resource constraints, more acute for small satellites, have resulted in an increased adoption of Commercial-Off-The-Shelf (COTS) parts in the design of satellites. While COTS parts must undergo a series of tests to become suitable for operation in the space environment (“space-rated”), their frequent use may still represent a challenge in terms of reliability. For example, “COTS manufacturers may implement a processing change resulting in a small performance impact in an Earth environment, but a serious impact under a space radiation environment” [19]. As a result, the “burden of proof [is] on the user and not the manufacturer”, which once again may exceed the testing capabilities of, or resources allocated to a small mission. Specifically, the radiation response of COTS devices can be “difficult to characterize due to large [variability] within a lot and the difficulty of testing imposed by packaging and hybridization” [20].

We hypothesize that the differences in reliability and failure behavior on-orbit between small and large satellites, especially with regards to infant mortality, may be due in part to a compounded problem of more reliance on COTS for the small satellites, and less testing and/or modifications of these parts to make them suitable for space environment. This hypothesis might address in part why satellites in the small satellite category exhibit the lowest reliability of the three satellite categories during the first 9 years on Fig. 2.

Factors intrinsically related to the design

The mass and geometric limitations (volume and other dimensions) imposed on small spacecraft can translate into several potential causes of failure behavior. For example:

---

[*] These broad categories of potential causes for the differences in observed failure behavior are not meant to be exhaustive.

- Small spacecraft, for which weight represents a critical parameter, cannot afford to have as much redundancy as larger spacecraft. They are therefore often based on “single-string” designs, which can turn a simple anomaly into a complete loss of the spacecraft.

- The design of small spacecraft relies on a greater package density, which can expose certain parts (such as plastic-encapsulated microcircuits or PEM) to higher temperatures, resulting in an increased risk of failure [6].

- “By its nature, a small spacecraft offers less natural shielding” than large spacecraft [6], and is therefore more exposed to the effects of cumulative radiation. For example, a form of radiation damage called displacement damage dose (DDD) “degrades the performance of solar cells, detectors, opto-couplers and optical lenses. It is more difficult to harden against DDD, therefore, the use of shielding […] is used to mitigate its effects” [20]. As a result, small satellites that do not benefit from sufficient shielding may be more subject to radiation-induced failures than larger spacecraft.

Factors that can potentially affect the failure behavior of large satellites include the following:

- Scaling up subsystems that are usually designed to operate within small/medium host spacecraft poses several challenges. Typically, larger and heavier spacecraft experience higher structural and electrical loads than smaller satellites. Power systems may generate excessive heat that accelerates the physical degradation of parts over time. This phenomenon may contribute to the increased wear out experienced by large satellites after 8 years, as seen in Fig. 2.

- The complexity of large satellites may also influence the infant mortality observed in this category of spacecraft. Large spacecraft with multiple instruments and subsystems require intensive wiring and an increased number of interfaces, adding potential failure points. As the number of connections between subsystems increases, the integration process becomes more delicate. This in turn may increase the likelihood of human errors, which can translate into a higher number of failures observed during the early life of large satellites.

Finally, we noted based on Fig. 2 that satellites within the Medium category exhibited the highest reliability profile of the three satellite categories, with the lowest infant mortality and a moderate wear-out behavior. These medium-sized satellites may not incur the penalties of small satellites discussed previously (budget restrictions, parts selection, residual fragility) and they are less likely to be subject to the challenges of larger satellites. In other words, medium-sized satellites may benefit from “the best of both worlds” (Small and Large categories) in terms of reliability. More importantly, they may simply correspond to the range of size for which parts, equipment, and design practices are currently the most mature and appropriate.

7. CONCLUSION

Reliability has long been recognized as a critical attribute for space systems, and potential causes of on-orbit failures are carefully sought for identification and elimination through various types of testing prior to launch. Several parameters or characteristics of the design, such as mission type, orbit, or spacecraft complexity, can potentially affect the probability of failure of satellites. In this paper, we explored the correlation between satellite mass, considered here as a proxy for size, and satellite reliability, and we investigated whether different classes of satellite, defined in terms of mass, exhibit different reliability profiles. To do so, we performed a statistical analysis on a sample of 1,444 Earth-orbiting satellites successfully launched between January 1990 and October 2008, by defining three main categories of satellites according to their mass: Small (0–500 kg), Medium (500–2,500 kg), and Large (>2,500 kg).

From the failure data, we first derived nonparametric reliability curves for each satellite mass category. We then conducted a parametric analysis by fitting the nonparametric results with Weibull distributions. Two-Weibull mixture distributions proved good candidates to accurately represent the nonparametric curves of reliability for each mass category. From the parametric results, we conducted a more detailed comparative analysis of the failure rate and failure behavior on orbit of the three satellite mass categories. Our results show that small satellites experience the highest infant mortality of all three satellite mass categories. Past one year and a half, we found that small and medium satellites exhibit a very similar (conditional) reliability behavior. Our results also identified distinct wear-our failure only for the large satellite category (after roughly 8 years on orbit).

Finally, we concluded this work by formulating a set of hypotheses that may address the causal factors for the statistical differences in the failure behavior of the three satellite categories. The possible causes for these differences were categorized and discussed under three headings: 1) differences in testing phase of the satellite prior to launch; 2) differences in to procurement and parts selections; and 3) differences intrinsically related to the design type (size of the spacecraft), such as increased structural and thermal loads in larger satellites which may contribute to wear-out failures.

We hope this work provides helpful feedback to the space industry for better understanding on orbit failure behavior of satellites, and ultimately for redesigning satellite test and screening programs, parts selection and redundancy allocation.

APPENDIX: CONFIDENCE INTERVAL ANALYSIS

The Kaplan-Meier estimator (Eq. 1) provides a maximum likelihood estimate of reliability but does not inform us about the dispersion around $\hat{R}(t_i)$. This dispersion is captured by the variance or standard deviation of the estimator, which is then used to derive the upper and lower bounds for say a 95% confidence interval (that is, a 95% likelihood that the actual reliability will fall between the two calculated bounds, with the Kaplan-Meier analysis providing us with the most likely estimate). The variance of the estimator is provided by Greenwood's formula (Eq. 4):

$$\hat{\mathrm{var}}\left[R(t_i)\right] \equiv \sigma^2(t_i) = \left[\hat{R}(t_i)\right]^2 \cdot \sum_{j \le i} \frac{m_j}{n_j(n_j - m_j)} \qquad (11)$$

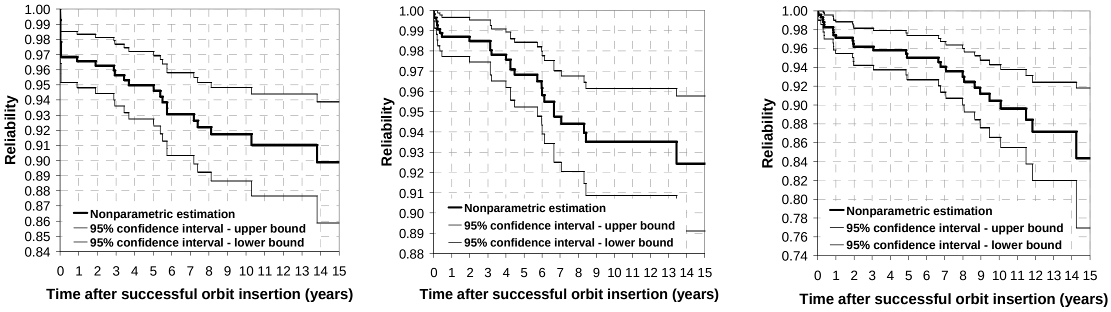
And the 95% confidence interval is determined by:

$$R_{95\%}(t_i) = \hat{R}(t_i) \pm 1.96 \cdot \sigma(t_i) \qquad (12)$$

More details about these equations can be found in [21, 22, 23].

When Eqs. (11) and (12) are applied to the data within each category along with the Kaplan-Meier estimated satellite reliability $\hat{R}(t_i)$ shown in Fig. 2, we obtain the 95% confidence interval curves. These results for each mass category of satellites are shown in Fig. 10.

Fig. 10 shows for example that the reliability of small satellites four years after orbit insertion will fall between 92.7% and 97.2% with a 95% likelihood (confidence interval). In addition, the most likely reliability estimate is at t = 4 years for the small satellites is the Kaplan-Meier result $\hat{R}(t = 4\,\text{years}) = 95.0\%$. Notice that the dispersion of $R(t_i)$ around $\hat{R}(t_i)$ increases with time. This increase in dispersion can be seen in Fig. 10 by the growing gap between the Kaplan–Meier estimated reliability and the confidence interval curves. This phenomenon illustrates the increasing uncertainty or loss of accuracy of the statistical analysis of satellite reliability with time resulting from the decreasing sample size (see discussion in Section 1 regarding the limitation of data specialization for satellite reliability analysis).



Small category Medium category Large category

Fig. 10. Satellite reliability with 95% confidence intervals for each mass category

REFERENCES

[1] J-F. Castet, J.H. Saleh, Satellite reliability: statistical data analysis and modeling, Journal of Spacecraft and Rockets, Vol. 46, No. 5, Sept-Oct. 2009.

[2] H. Hecht, Reliability for Space Mission Planning, in: J.R. Wertz, W.J. Larson, Space Mission Analysis and Design, 3rd ed., Microcosm Press, Hawthorne, CA, and Springer, New York, NY, 1999, pp. 777-778.

[3] M. Williamson, Spacecraft Technology: The Early Years, Institute of Electrical Engineers, Herts, United Kingdom, 2006, p. 306.

[4] D.A. Bearden, A complexity-based risk assessment of low-cost planetary missions: when is a mission too fast and too cheap?, Acta Astronautica 52 (2003) 371–379.

[5] R. Fleeter, Design of Low-Cost Spacecraft in: J.R. Wertz, W.J. Larson, Space Mission Analysis and Design, 3rd ed., Microcosm Press, Hawthorne, CA, and Springer, New York, NY, 1999, pp. 853-863.

[6] L.P. Sarsfield, The Cosmos on a Shoestring – Small Spacecraft for Space and Earth Science, RAND, Santa Monica, CA, 1998.

[7] Technology for Small Spacecraft, Panel on Small Spacecraft Technology, National Research Council, National Academy Press, Washington D.C., 1994.

[8] ANSI/AIAA Guide for Estimating and Budgeting Weight and Power Contingencies for Spacecraft Systems G-020-1992.

[9] E.L. Kaplan, P. Meier, Nonparametric Estimation from Incomplete Observations, Journal of the American Statistical Association 53(282) (1958) 457–481.

[10] J-F. Castet, J.H. Saleh, Satellite and satellite subsystems reliability: statistical data analysis and modeling, Reliability Engineering and System Safety, Vol. 94, No. 11, pp. 1718-1728, Nov. 2009.

[11] J-F. Castet, J.H. Saleh, Weibull Modeling for Nonparametric Satellite Reliability: Graphical Versus Maximum Likelihood Estimations, submitted to Journal of Spacecraft and Rockets, April 2009.

[12] J.F. Lawless, Statistical models and methods for lifetime data, 2nd ed., John Wiley & Sons, New York, 2003.

[13] J-F. Castet, J.H. Saleh, Single versus mixture Weibull distributions for nonparametric satellite reliability, submitted to Reliability Engineering and System Safety, May 2009.

[14] C.E. Ebeling, An Introduction to Reliability and Maintainability Engineering, McGraw-Hill, New York, 1996.

[15] D. Kececioglu, Reliability Engineering Handbook, 1st ed., Prentice-Hall, Englewood Cliffs, NJ, 1991.

[16] L.M. Leemis, and M. Beneke, Burn-In Models and Methods: A Review, IIE Transactions 22 (2) (1990), pp. 172–180.

[17] H.W. Block, and T.H. Savits, Burn-In, Statistical Science, 12(1), (1997), p. 1.

[18] T.E. Gindorf, R.F. Miles Jr., G.B. Murphy, Space-hardware design for long life with high reliability, Proceedings of the Annual Reliability and Maintainability Symposium, 1994, pp. 338–341.

[19] S. Kayali, Utilization of COTS electronics in space application, reliability challenges and reality, Commercialization of Military and Space Electronics Conference & Exhibition, Los Angeles, CA, 12-13 Feb. 2002.

[20] J.L. Barth, Prevention of Spacecraft Anomalies – The Role of Space Climate and Space Weather models, in: I.A. Daglis, Effects of Space Weather on Technology Infrastructure, Springer Netherlands, 2005, pp. 125–128.

[21] J.I. Ansell, M.J. Phillips, Practical methods for reliability data analysis, Clarendon Press, Oxford, NY, 1994.

[22] W.O. Meeker, L.A. Escobar, Statistical methods for reliability data, John Wiley & Sons, New York, NY, 1998.

[23] M. Rausand, A. Høyland, System reliability theory: models, statistical methods, and applications, 2nd ed., Wiley-Interscience, Hoboken, NJ, 2004.