Software Failure and Reliability Assessment Tool: Report

Author Name

2019-06-08_21:01

Tab 1: Select, Apply, and Analyze Data

Sample of the updated data ('CSR2') in different formats:

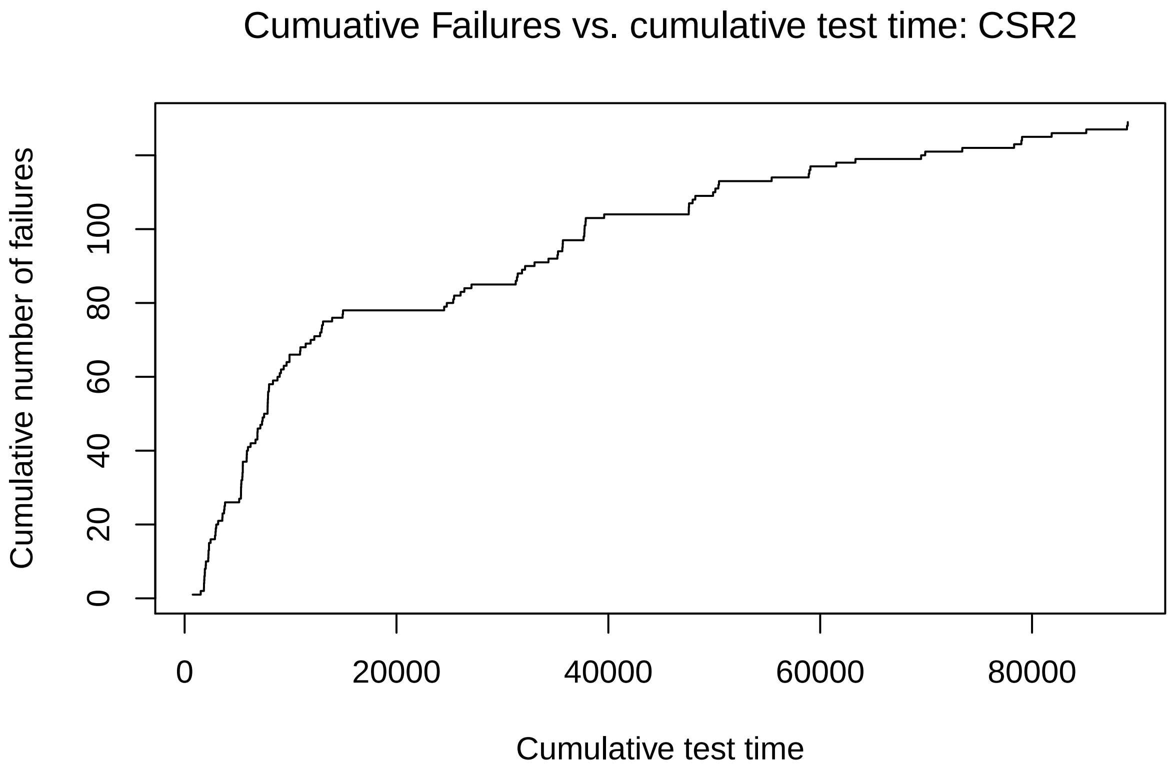
The table below shows the first ten points of the input data CSR2. 'FC', 'CFC', 'FT', 'IF', and 'FN' indicates failure counts, cumulative failure counts, failure times, interfailure times, and number of failures respectively.

Table 1: First ten points of the input data

FN	IF	FT
1	760	760
2	758	1518
3	303	1821
4	6	1827
5	22	1849
6	14	1863
7	42	1905
8	4	1909
9	84	1993
10	15	2008

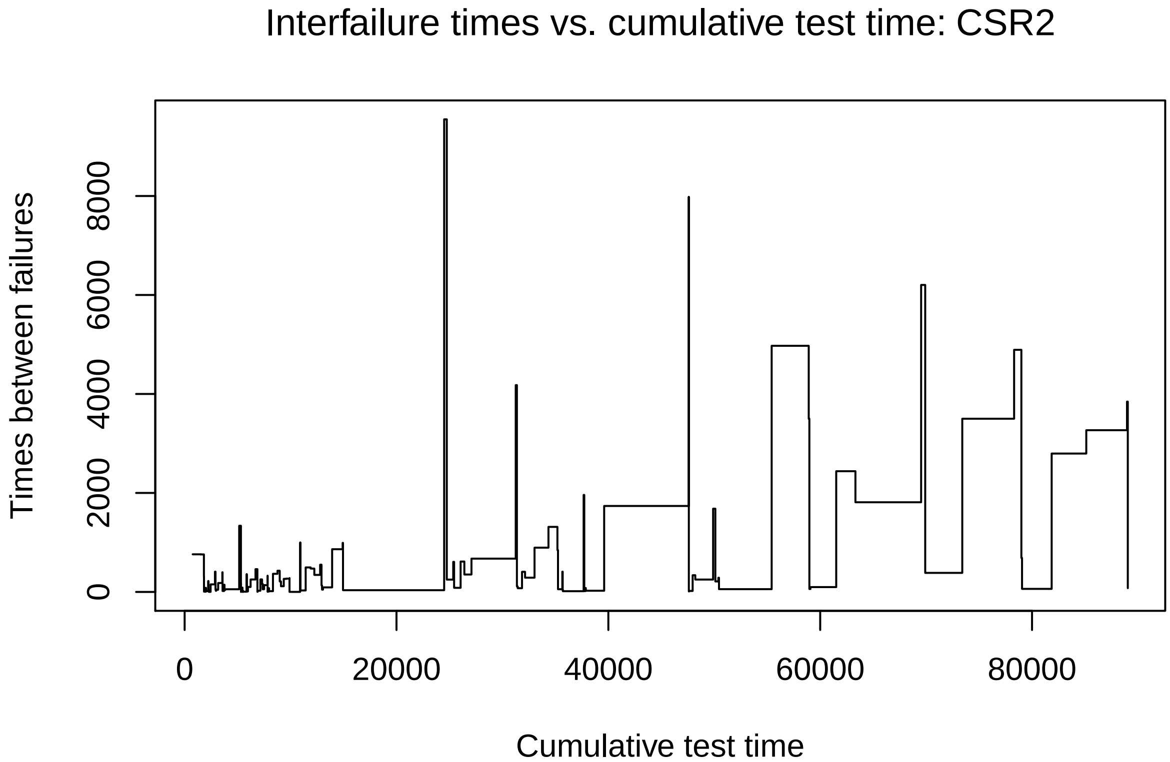
Cumulative failures

The following figure shows the CSR2 data as the cumulative number of failures (FN) detected as a function of cumulative test time (FT). An increasing trend indicates periods where more faults were detected. Ideally, the cumulative number of failures should level off to a horizontal line, indicating that no new faults have been detected.



Times between failures/Interfailure times

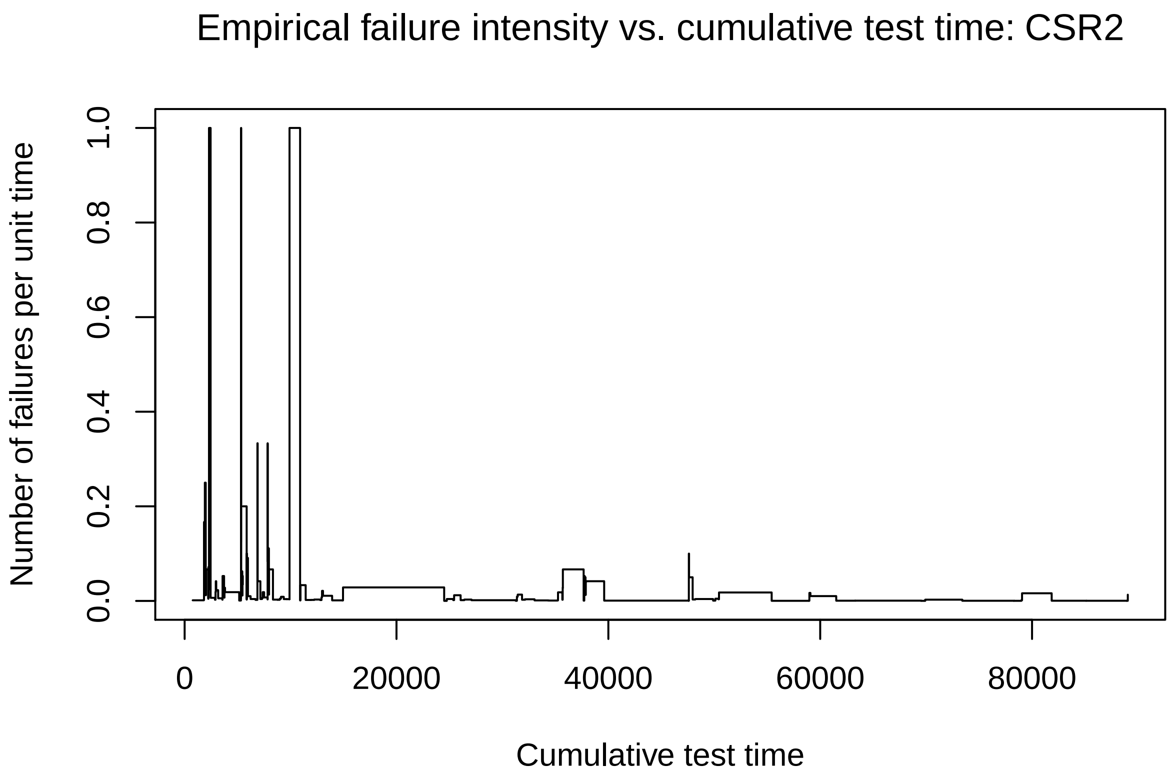
The following figure shows the CSR2 times between failures (IF) as a function of cumulative test time (FT). An increasing trend indicates periods where fewer faults were detected. Ideally, the time between failures should increase, indicating that no new faults have been detected.



Failure intensity

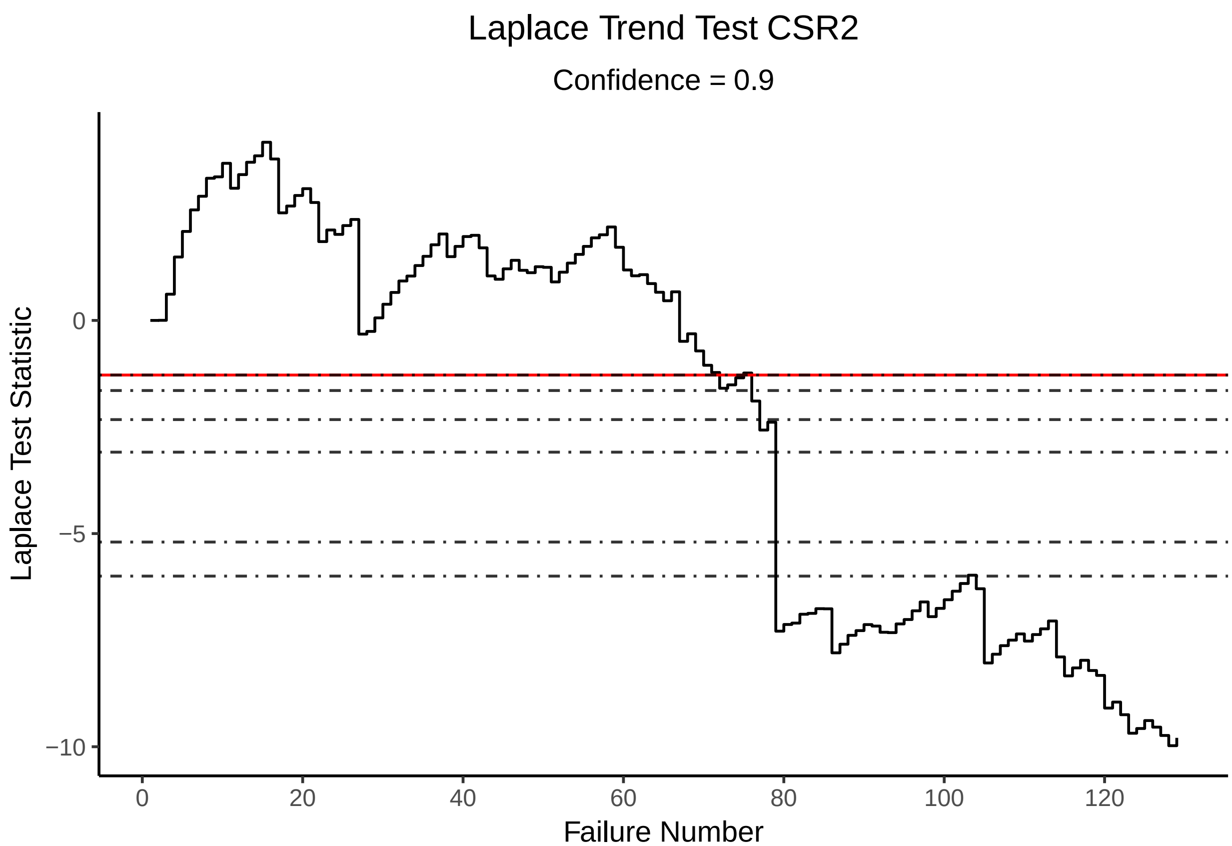
The following figure shows the CSR2 data as the number of failures detected per unit time as a function of cumulative test time (FT). A decreasing trend indicates periods where fewer faults were detected. Ideally, the failure intensity should decrease, indicating that no new faults have been detected.

A decrease in failure intensity indicates increase in reliability of the software subjected to testing. Ideally, the failure intensity should go to zero.



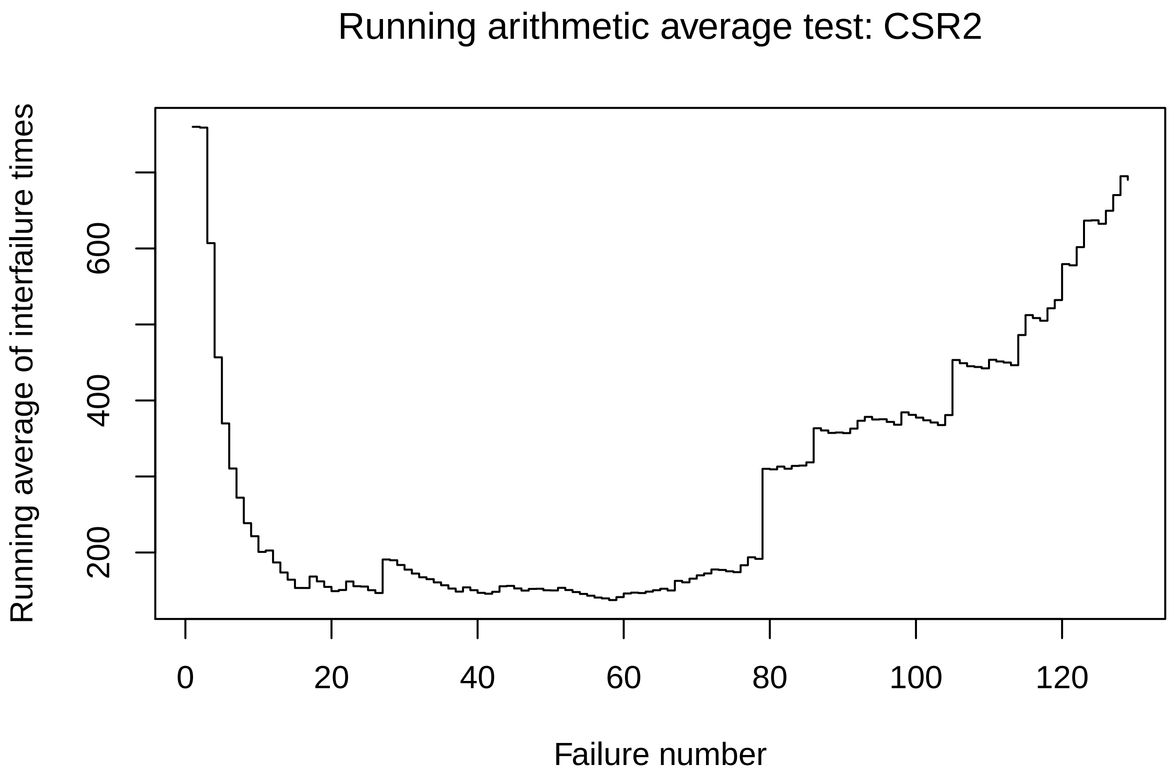
Laplace Trend Test

The following figure shows the Laplace test statistic for reliability growth as a function of cumulative test time (FT). A decreasing trend indicates reliability growth, while an increasing trend indicates reliabilty deterioration. The Laplace test statistic on the y-axis corresponds to the critical values of a normal distribution. This means that if the trend falls below a specific level, then we cannot reject the null hypothesis that the failure data suggests reliability growth at a specified level of confidence. The six black dot-dash style lines correspond to the 90%, 95%, 99%, 99.9%, 99.9999%, and 99.999999% respectively. The red line is user-specified and has been set to 90%. The level of confidence is a subjective choice made by the analyst. Reliability growth is desired because software reliability growth models assume curves that exhibit increasing time between failures. If reliability growth is not present than the model fitting step may fail or produce predictions that are inaccurate. Therefore, the Laplace test statistic provides an objective quantitative measure for the analyst to decide if predictions may or may not be accurate.



Running arithmetic average

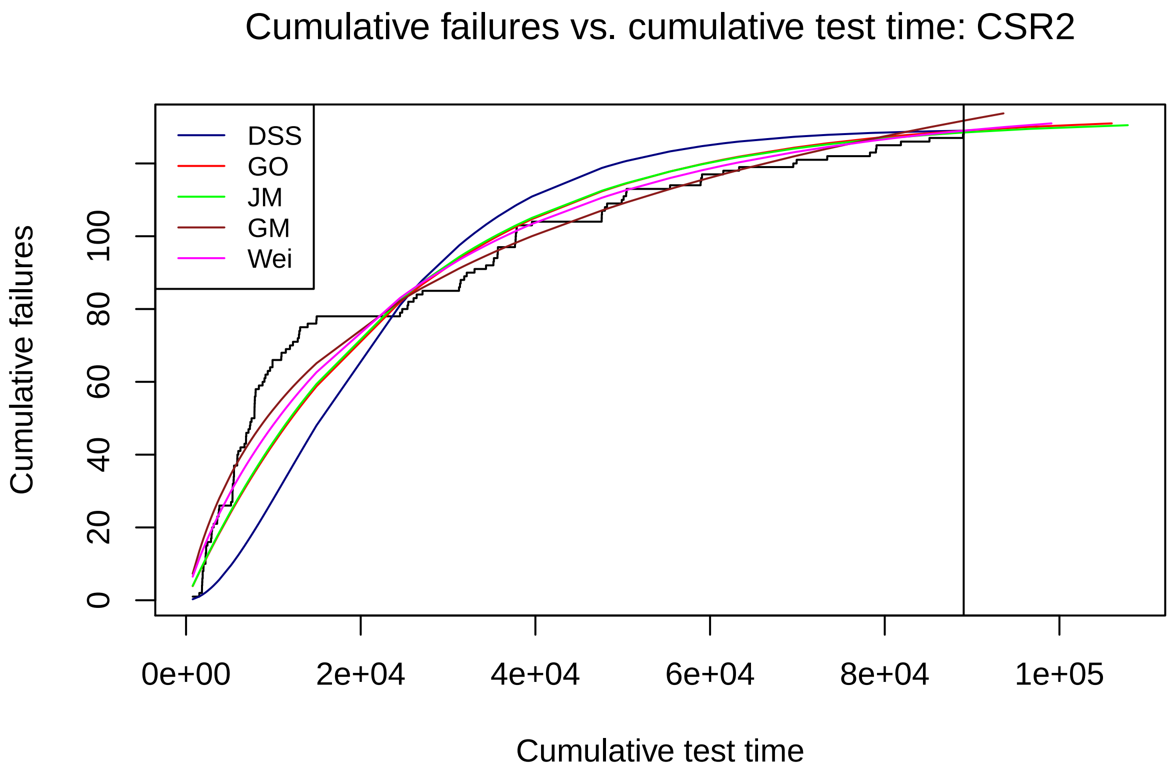
The running arithmetic average plots the average of the first k times between failure. An increasing trends indicates reliability growth, while a decreasing trend indicates reliability deterioration. This is intuitive because if the time between failures is increasing, then later failures will increase the average.



Tab2: Set Up and Apply Models

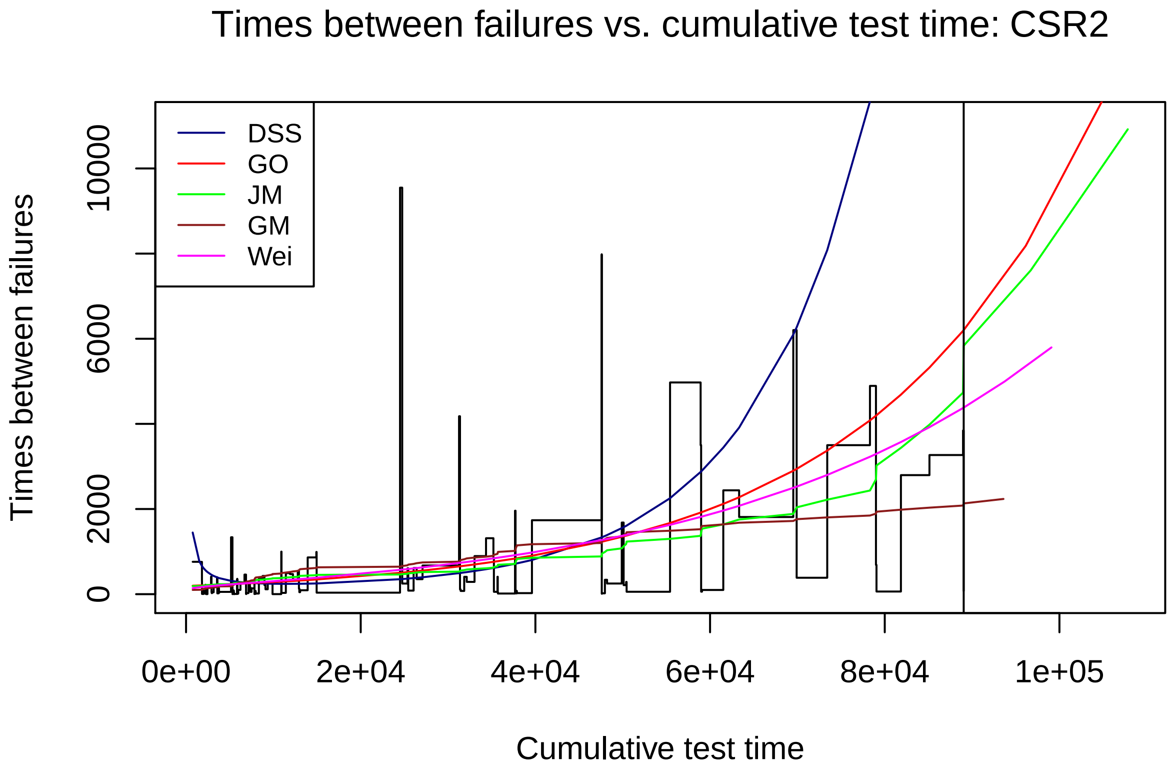
Cumulative failures

The following figure shows the fit of delayed s-shaped, geometric, Weibull, Goel-Okumoto, Jelinski-Moranda models to the cumulative number of failures detected in the CSR2 data.



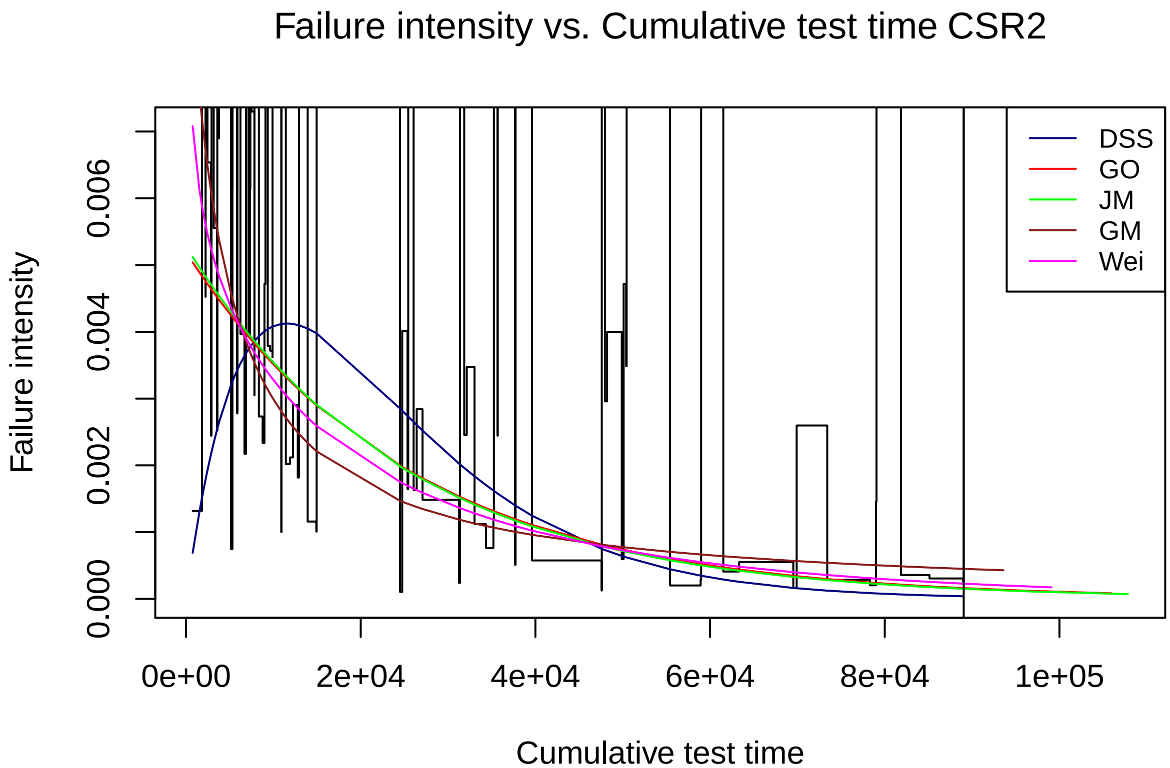
Times between failures

The following figure shows the fit of delayed s-shaped, geometric, Weibull, Goel-Okumoto, Jelinski-Moranda models to the times between failures detected in the CSR2 data.



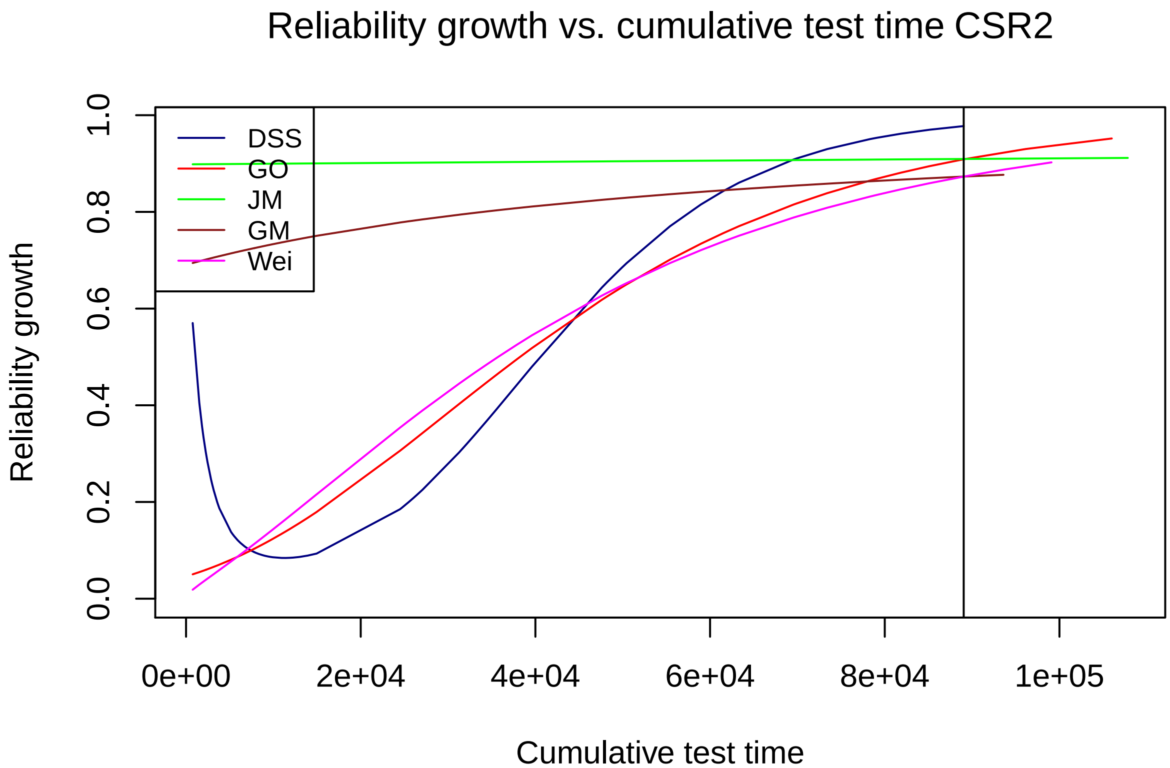
Failure intensity

The following figure shows the fit of delayed s-shaped, geometric, Weibull, Goel-Okumoto, Jelinski-Moranda models to the failure intensity of the CSR2 data.



Reliability growth

The following figure shows the reliability growth curve of the fit of delayed s-shaped, geometric, Weibull, Goel-Okumoto, Jelinski-Moranda models to the CSR2 data. The data itself does not display. This plot indicates a models prediction that the software will be reliable (exhibit zero failures) for a duration of 600 time units as a function of cumulative test time (FT). Selecting a model upon which to base a reliability assessment is a subjective choice made by the analyst. Statistical measures of goodness of fit, reported on page 12 of this report can be used to decide this decision making process. If the Laplace test statistic does not exhibit reliability growth, than a conservative approach is to document this as the reason why no reliability estimate is provided at the time of preparing a report.



Tab3: Query Model Results

The following table shows inferences enabled by the models, including the time to achieve a reliability of 90% (probability of zero failures for 600 time units), expected number of failures in the next 4116 time units, and expected time to observe an additional 3 failures computed for the fit of delayed s-shaped (DSS), geometric (GM), Weibull (Wei), Goel-Okumoto (GO), Jelinski-Moranda (JM) models to the models.

	Time to achieve specified reliability	Expected number of failures	Expected time to N failure
DSS	R = 0.9 achieved	0.137258900924792	NA
GO	R = 0.9 achieved	0.613094200899809	33159.9669281473
JM	84926.453285578	0.574204486242337	39260.0753318027
GM	145038.680564181	1.81206384825529	6912.65185177134
Wei	9113.09030502735	0.887137365337594	16363.1270434606

Tab4: Evaluate Models

The following table shows the measures of goodness of fit computed for the delayed s-shaped, geometric, Weibull, Goel-Okumoto, Jelinski-Moranda The Akaike Information Criterion (AIC) is an information theoretic measure. Lower values are preferred. The GM model achieved the lowest AIC value on the CSR2 data. A difference of 2.0 or more in the AIC values of two models indicates the model with the lower AIC score is preferred with statistical significance. The Predictive Sum of Squares Error (PSSE) used 90% of the CSR2 data to fit the models and computed the sum of the squares between the differences of the remaining 10% of the data not used to fit the models. Lower values are preferred. The GM model achieved the lowest PSSE value on the CSR2 data. The measures of goodness of fit can help select a model, but the choice is ultimately a subjective choice made by the analyst.

	Akaike Information Criterion (AIC)	"Predictive sum of squares error (PSSE)" ~ 0.9
DSS	1913.69	364.94
GO	1848.57	101.54
JM	1845.03	84.79
GM	*1838.86	74.41
Wei	1846.59	*45.98