

# Software Failure and Reliability Assessment Tool: Report

*xxx*

*2018-09-10\_19:44*

## **Tab 1: Select, Apply, and Analyze Data**

### **Sample of the updated data ('SYS1') in different formats:**

The table below shows the first ten points of the input data SYS1. 'FC', 'CFC', 'FT', 'IF', and 'FN' indicates failure counts, cumulative failure counts, failure times, interfailure times, and number of failures respectively.

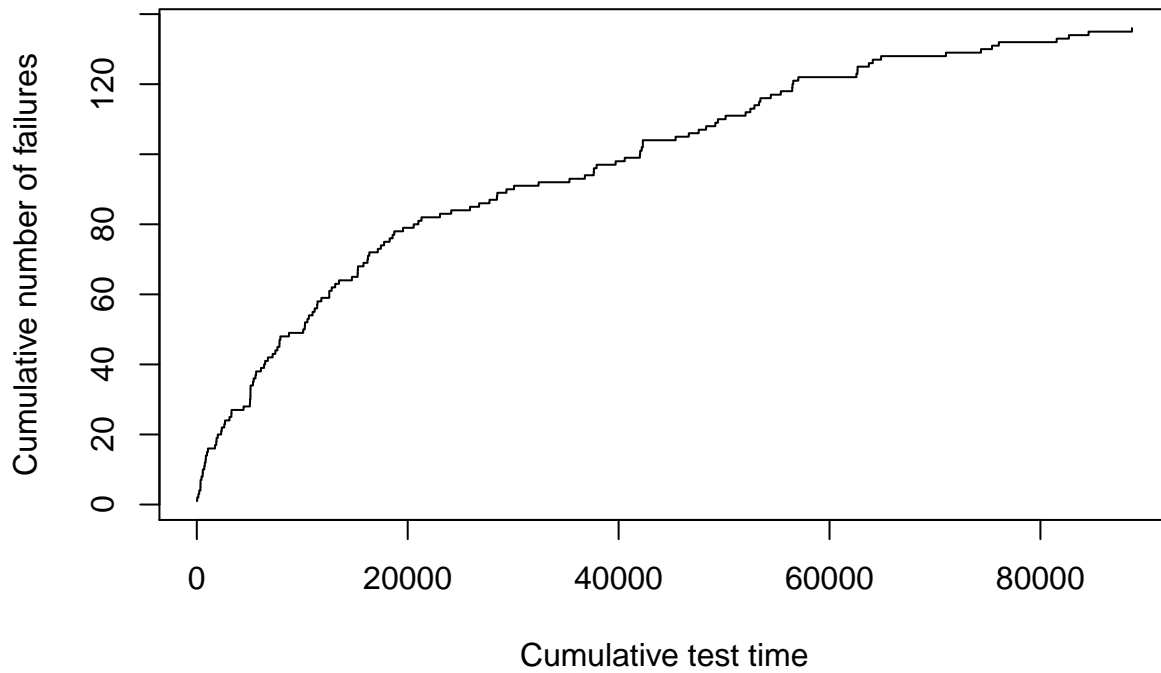
Table 1: First ten points of the input data

FN	IF	FT
1	3	3
2	30	33
3	113	146
4	81	227
5	115	342
6	9	351
7	2	353
8	91	444
9	112	556
10	15	571

## Cumulative failures

The following figure shows the SYS1 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 failure should level off to a horizontal line, indicating that no new faults have been detected.

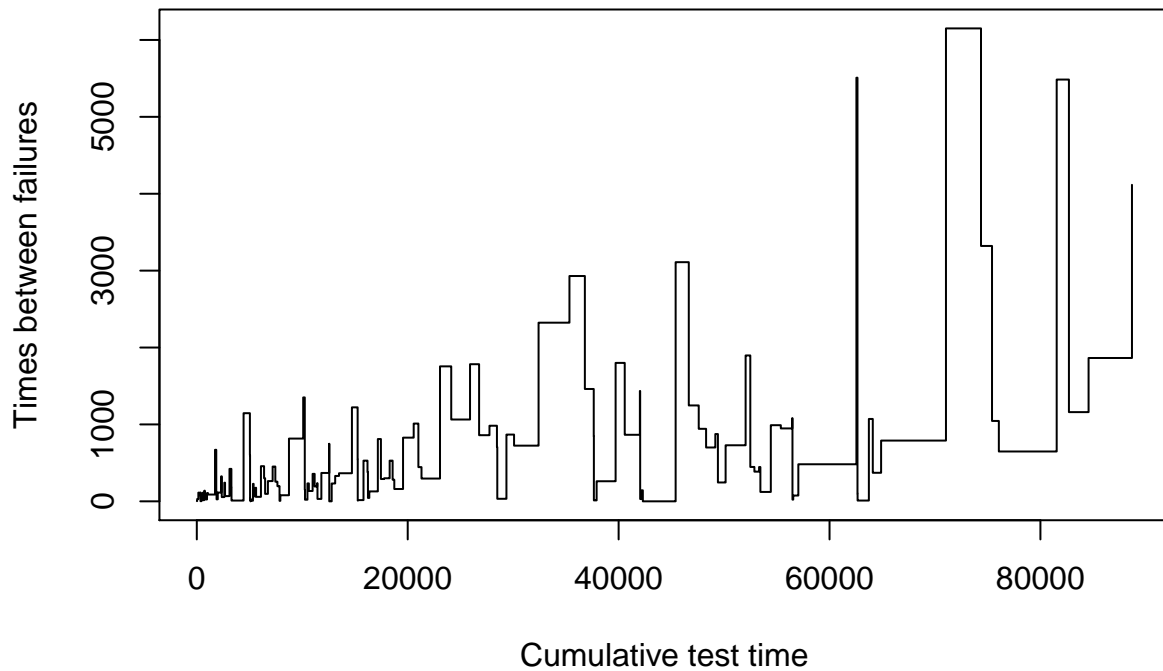
Cumulative Failures vs. cumulative test time: SYS1



## Times between failures/Interfailure times

The following figure shows the SYS1 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.

Interfailure times vs. cumulative test time: SYS1

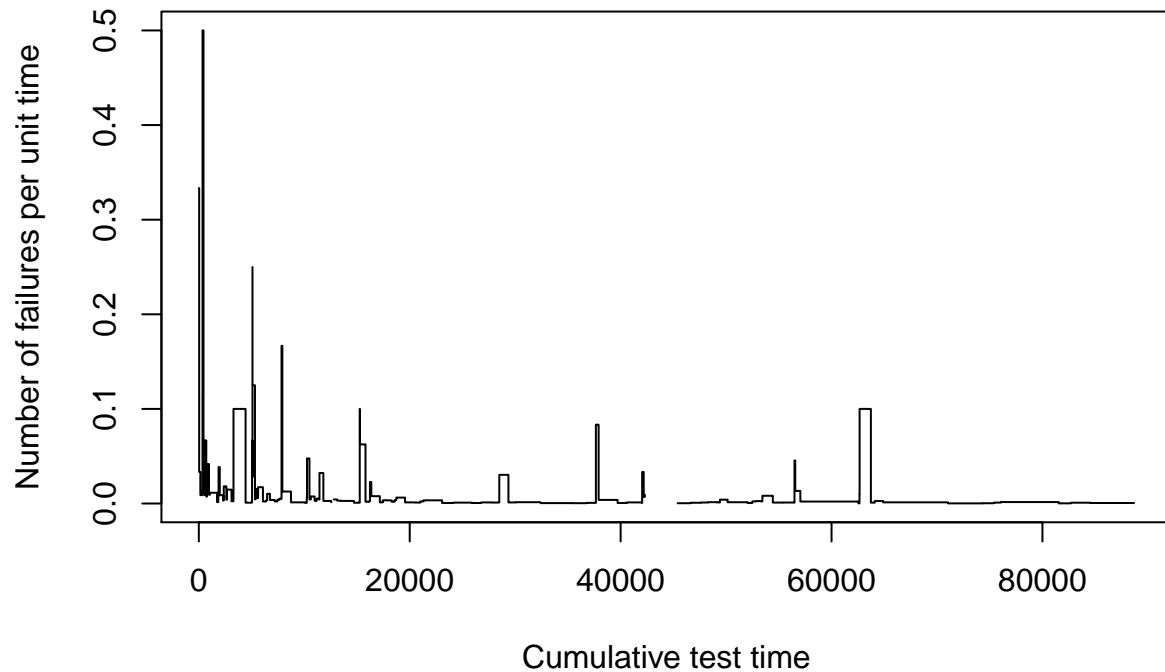


## Failure intensity

The following figure shows the SYS1 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.

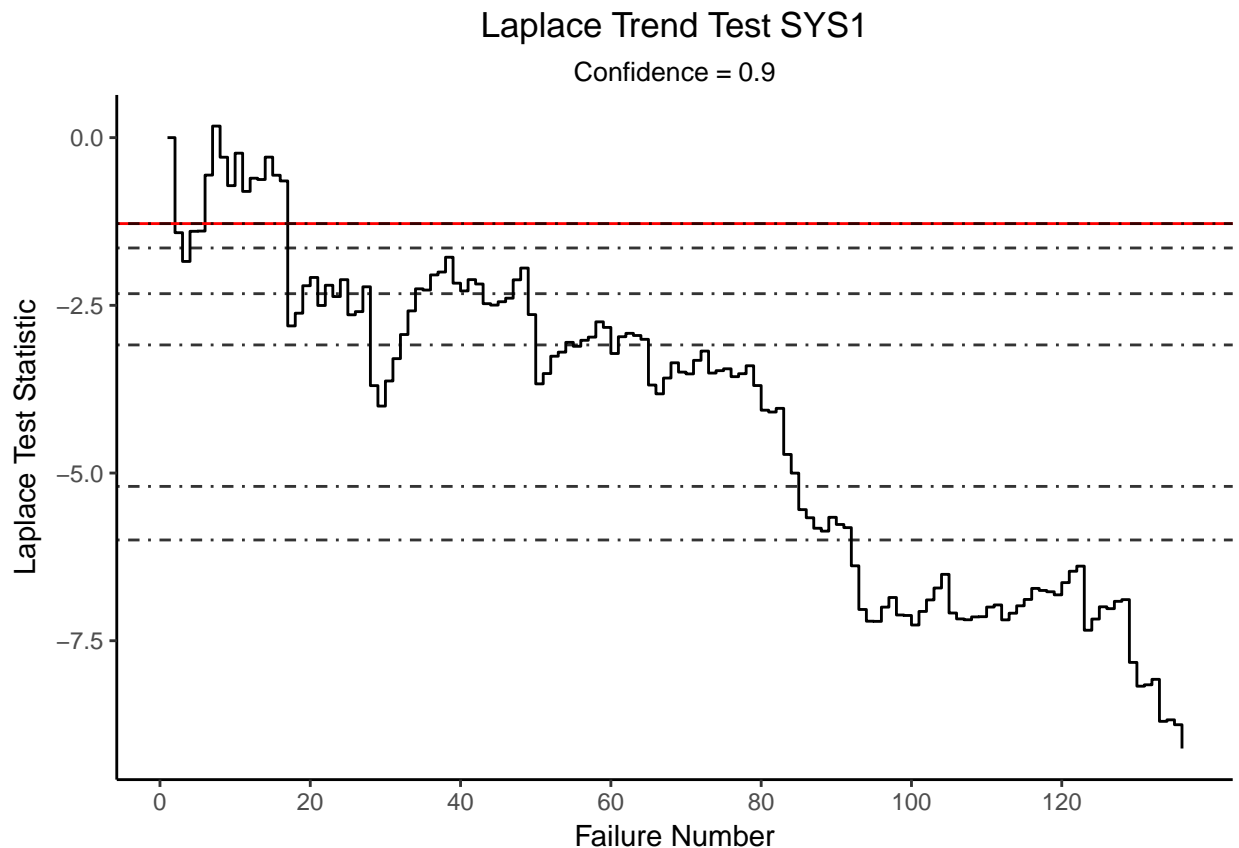
Decrease in failure intensity indicates increase in reliability of the software subjected to testing. Ideally, failure intensity should go to zero.

Empirical failure intensity vs. cumulative test time: SYS1



## Laplace Trend Test

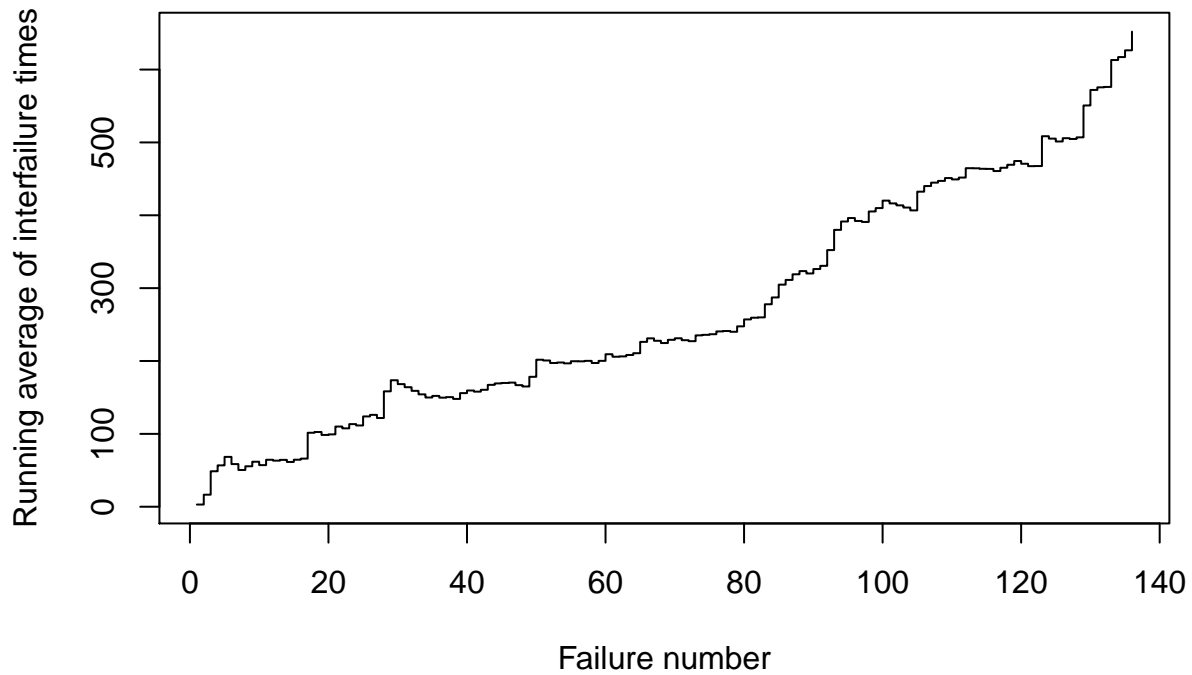
The following figure shows the Laplace test statistic for reliability growth as a function of cumulative test time (FT). A decreasing trends indicates reliability growth, while an increasing trend indicates reliability deterioration. The Laplace test statistic on the y-axis corresponds to the critical 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 use-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 as a function of the  $k$ th failure. An increasing trend 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.

### Running arithmetic average test: SYS1



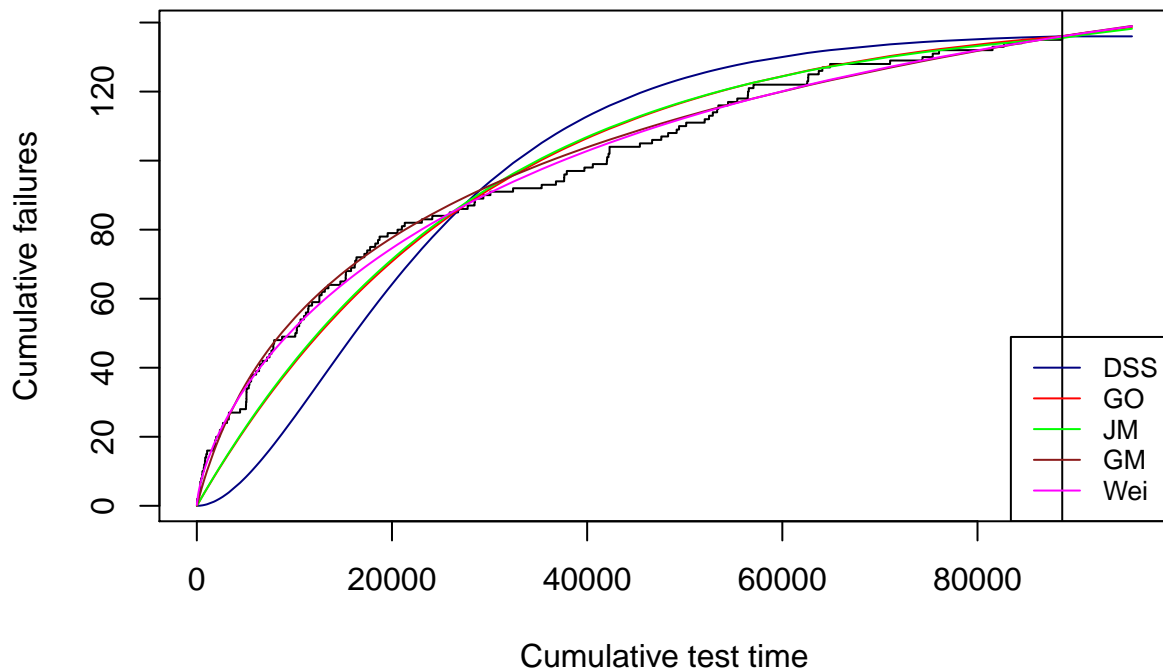
## Tab2: Set Up and Apply Models

### Cumulative failures

The following figure shows the DSS, GM, Wei, GO, JM models fit to the cumulative number of failures detected in the SYS1 data.

```
## [1] 3.847e-06 4.647e-04 9.042e-03 2.177e-02 4.911e-02 5.170e-02 5.229e-02
## [8] 8.232e-02 1.283e-01 1.353e-01 2.070e-01 2.366e-01 2.859e-01 3.022e-01
## [15] 3.807e-01 4.510e-01 1.163e+00 1.323e+00 1.358e+00 1.520e+00 2.023e+00
## [22] 2.115e+00 2.537e+00 2.662e+00 3.491e+00 3.872e+00 3.894e+00 6.677e+00
## [29] 8.347e+00 8.390e+00 8.495e+00 8.507e+00 8.507e+00 8.530e+00 9.200e+00
## [36] 9.395e+00 9.929e+00 1.011e+01 1.155e+01 1.252e+01 1.285e+01 1.373e+01
## [43] 1.528e+01 1.617e+01 1.687e+01 1.757e+01 1.759e+01 1.787e+01 2.088e+01
## [50] 2.605e+01 2.662e+01 2.671e+01 2.762e+01 2.814e+01 2.955e+01 3.031e+01
## [57] 3.124e+01 3.136e+01 3.283e+01 3.580e+01 3.580e+01 3.673e+01 3.804e+01
## [64] 3.949e+01 4.433e+01 4.646e+01 4.649e+01 4.656e+01 4.862e+01 5.008e+01
## [71] 5.025e+01 5.075e+01 5.383e+01 5.493e+01 5.605e+01 5.801e+01 5.905e+01
## [78] 5.963e+01 6.262e+01 6.617e+01 6.770e+01 6.871e+01 7.448e+01 7.781e+01
## [85] 8.309e+01 8.551e+01 8.817e+01 9.001e+01 9.010e+01 9.228e+01 9.404e+01
## [92] 9.930e+01 1.051e+02 1.078e+02 1.092e+02 1.092e+02 1.096e+02 1.124e+02
## [99] 1.136e+02 1.156e+02 1.156e+02 1.158e+02 1.159e+02 1.159e+02 1.196e+02
## [106] 1.209e+02 1.218e+02 1.225e+02 1.232e+02 1.234e+02 1.241e+02 1.255e+02
## [113] 1.258e+02 1.261e+02 1.264e+02 1.265e+02 1.271e+02 1.277e+02 1.283e+02
## [120] 1.284e+02 1.284e+02 1.287e+02 1.312e+02 1.312e+02 1.312e+02 1.316e+02
## [127] 1.317e+02 1.320e+02 1.337e+02 1.343e+02 1.345e+02 1.346e+02 1.354e+02
## [134] 1.355e+02 1.357e+02 1.360e+02 1.360e+02 1.360e+02
## [1] 138
```

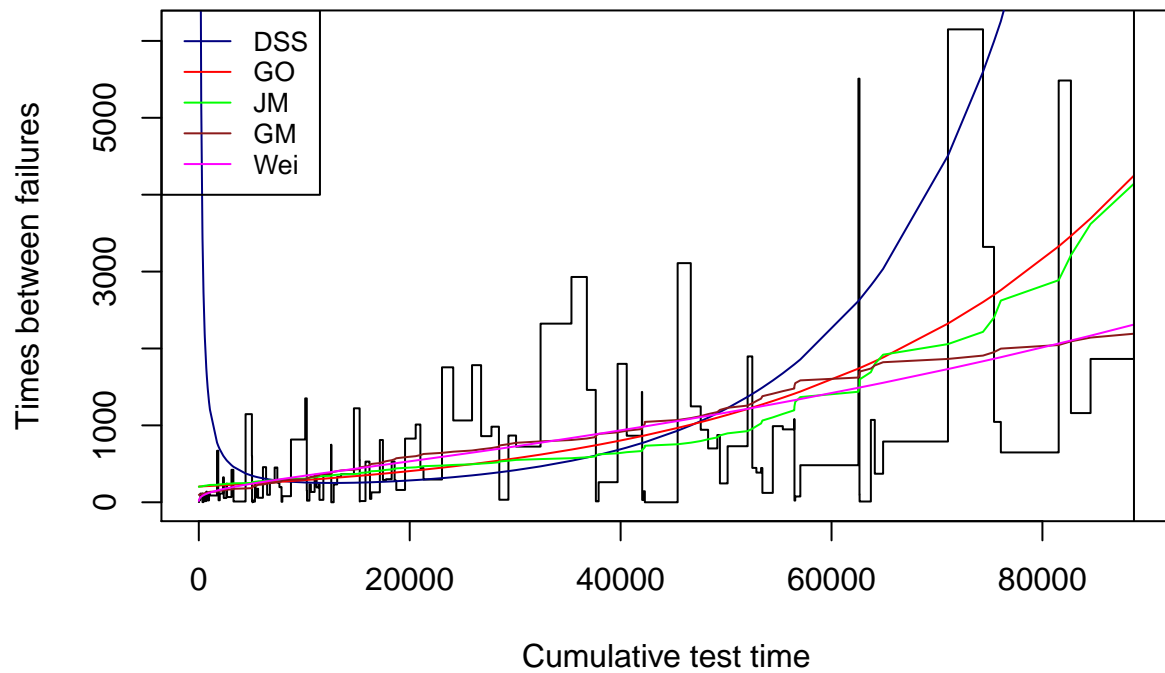
Cumulative failures vs. cumulative test time: SYS1



## Times between failures

The following figure shows the DSS, GM, Wei, GO, JM models fit to the times between failures detected in the SYS1 data.

Times between failures vs. cumulative test time SYS1

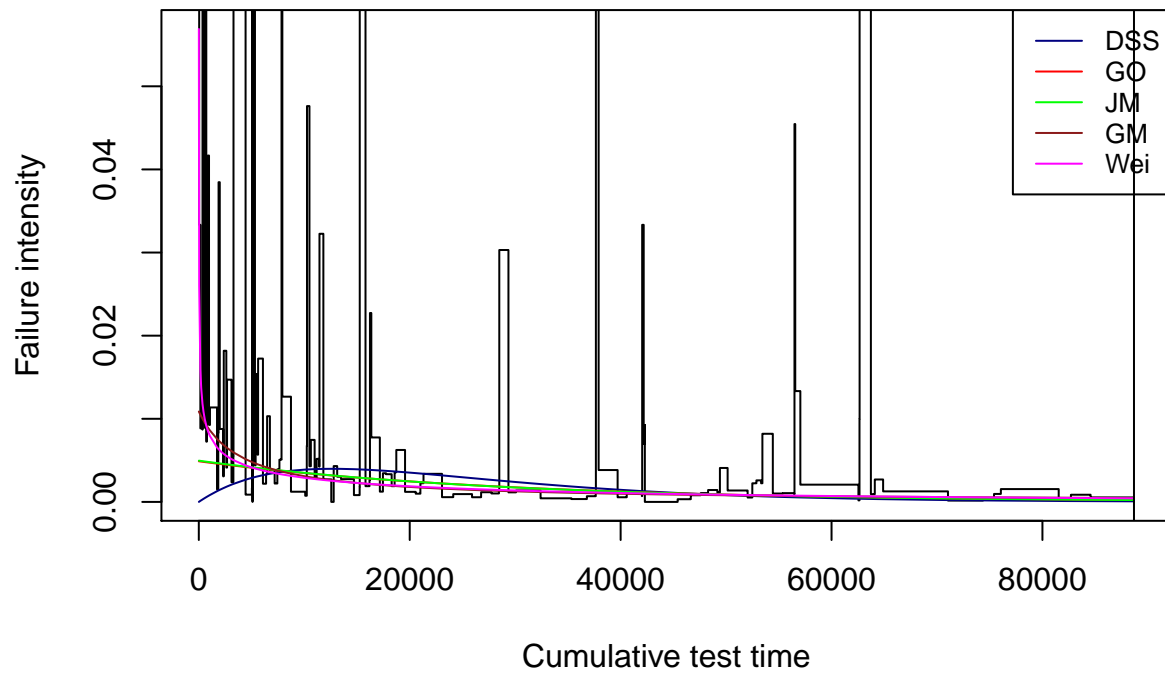




## Failure intensity

The following figure shows the DSS, GM, Wei, GO, JM models fit to the failure intensity of the SYS1 data.

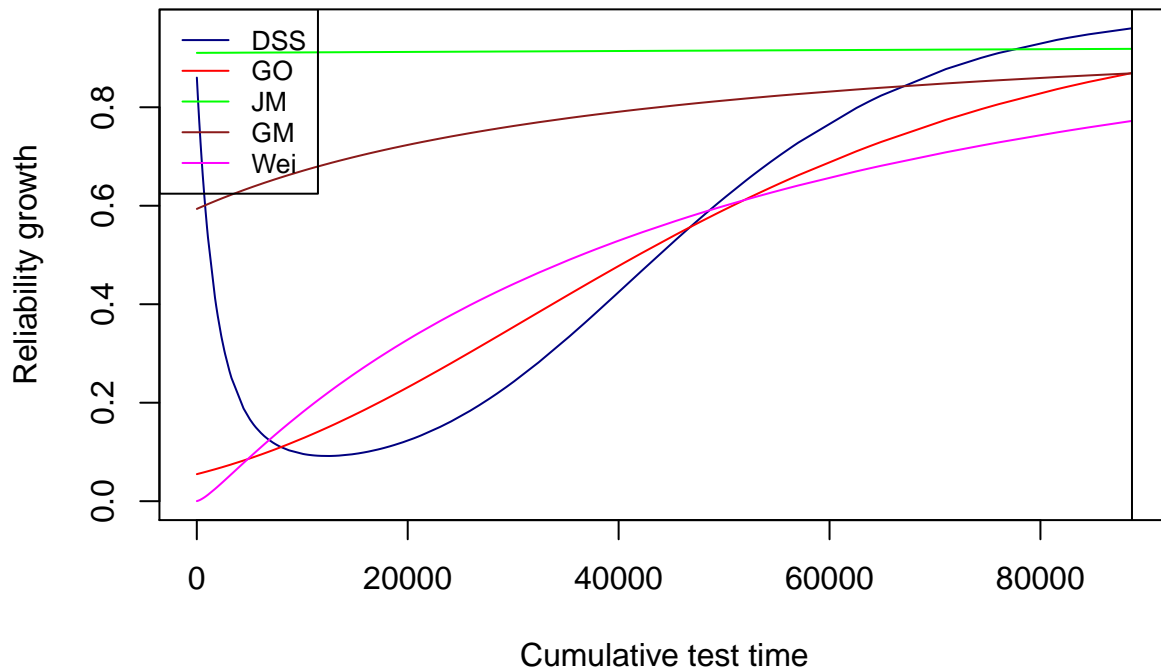
Failure intensity vs. Cumulative test time SYS1



## Reliability growth

The following figure shows the reliability growth curve of the DSS, GM, Wei, GO, JM models fit to the SYS1 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 xx 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.

Reliability growth vs. cumulative test time SYS1



### Tab3: Query Model Results

The following table shows inferences enabled by the models, including the time to achieve a reliability of xx (probability of zero failures for xx time units), expected number of failures in the next xx time units, and expected time to observe an additional xx failures computed for the Delayed s-Shaped (DSS), Geometric (GM), Weibull (Wei), Goel-Okumoto (GO), and Jelinski-Moranda (JM) models.

	Time to achieve specified reliability	Expected number of failures	Expected time to N failure
DSS	R = 0.9 achieved	0.246856262199799	NA
GO	8263.13681952821	0.903615409906593	10040.3808618466
JM	91142.2377161945	0.85612548252314	10742.5403828383
GM	153028.269493869	1.87747308675807	4390.88656760035
Wei	66732.9968495319	1.72595369956707	4793.97740413222

## Tab4: Evaluate Models

The following table shows the measures of goodness of fit computed for the DSS, GM, Wei, GO, JM. The Akaike Information Criterion (AIC) is an information theoretic measure. Lower values are preferred. The GM model achieved the lowest AIC value on the SYS1 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 SYS1 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 SYS1 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	2075.15	296.35
GO	1953.61	23.07
JM	1950.53	*19.6
GM	*1937.03	84.33
Wei	1938.16	74.94