

# Addressing potential reasons for discrepancies between Weibull Plots created by different organizations

Jason Freels

12/5/2021

## Background

In the B-1 eRCM meeting on Friday, 3 December, the RSO presented several Weibull plots that were produced using an algorithm developed by C3.ai. It's anticipated that this algorithm will be implemented as part of their PANDA dashboard. During this meeting some concerns were raised about differences between the reliability thresholds suggested by the Weibull plots created by C3.ai and the plots that were developed by PavCon.

PANDA is set to become the system of record for setting CBM+ reliability thresholds. The thresholds are based on the analyses presented in these Weibull plots. Therefore, it's critical that the methods implemented to create these plots be well documented.

This document highlights some possible reasons for discrepancies between the plots created by C3.ai and PavCon and suggests some possible ways to minimize them. It must be noted that the data used to create these Weibull plots can be very messy (more on this below). Consequently, choosing the **most correct** way to generate these plots involves answering two questions:

1. What is the analyst's approach to modeling the data?
2. How are the failure events defined and how they are included in the analysis?

These questions are fleshed out below. **Note:** I do not address the most basic and obvious question: "Are we sure that both organizations are using the same data?" Everything that follows assumes that they are.

## Question 1: What is the analyst's approach to modeling the data?

### Question 1(a) What estimator is used in these plots?

Figure 1 shows a Weibull plot I created from 150 randomly generated time to failure observations. This data set is composed of 17 observations from a  $WEIB(\beta = 4.2, \eta = 5.7)$  distribution and 133 observations from a  $WEIB(\beta = 0.943, \eta = 1754)$  distribution. This mixture of Weibull distributions was done to generate a data set that produces a Weibull plot that resembles those that have been presented for CBM+ fielded components. Suppose that all times are in hours.

To generate this - or any - Weibull plot, the  $x$  and  $y$  axes must be transformed in a specific manner (I won't go into this here). This is done because the first question we want to answer with a Weibull plot is - "Is the Weibull distribution a good model to describe the failure process that is revealed by this data". If the points fall **roughly** on a straight line when plotted on these transformed axes, we have some evidence that the Weibull distribution **is** a good choice.

But, what exactly are these points being plotted? The value of each point along the x-axis is pretty clear, these are the failure times (sorted from smallest to largest). The value of each point along the y-axis is more complicated. To determine the  $y$  value for each point, the analyst first has to choose an estimator for

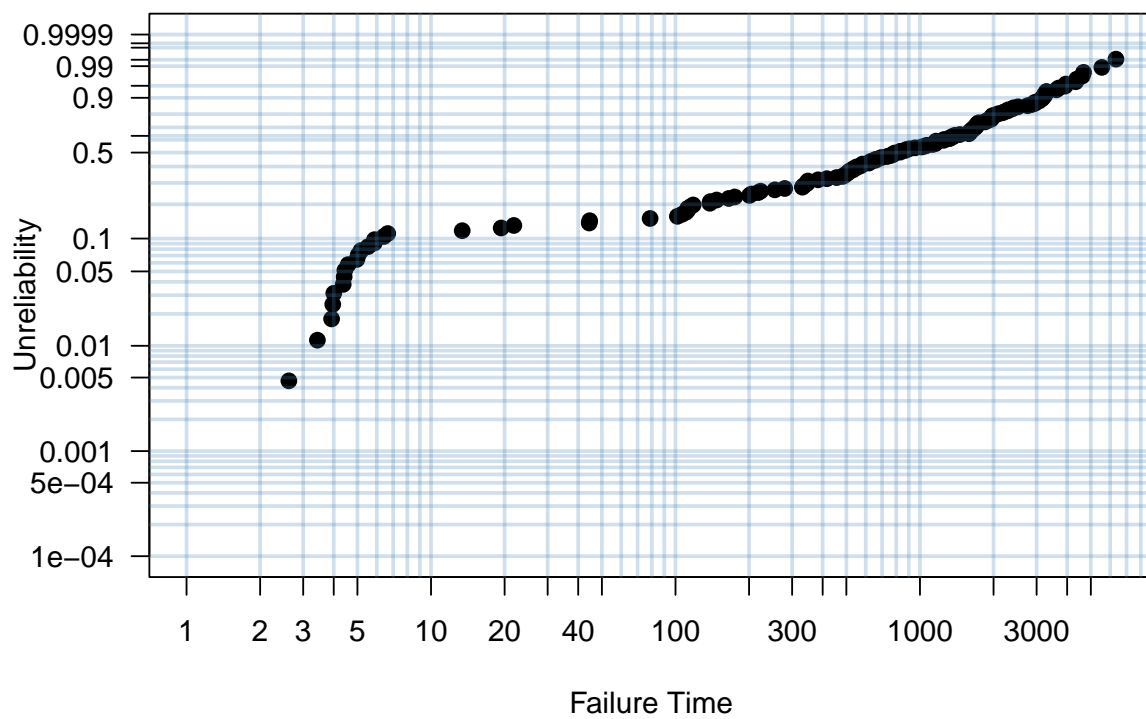


Figure 1: Randomly generated failures plotted on Weibull paper

the failure probability (aka Unreliability). Several estimators exist for this purpose, but some of the most commonly used are: [median-ranks](#), [Kaplan-Meier](#), and [Turnbull](#).

The plot shown in Figure 1 used the median ranks estimator. Figure 2, shown below, includes both the median-ranks estimate and the Kaplan-Meier estimate (the Kaplan-Meier estimate is shown in red). For this data set the difference between the estimates is small, and would likely not impact values of the Weibull parameters ( $\beta$ , and  $\eta$ ). For some data sets, however, the difference may be more pronounced.

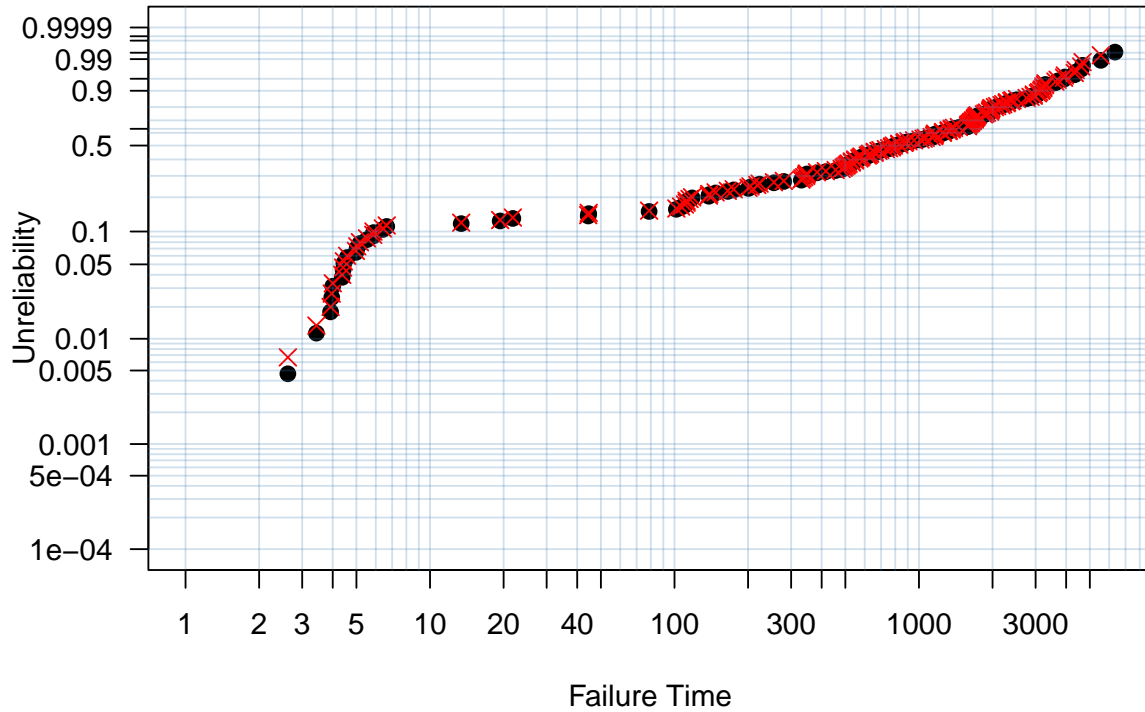


Figure 2: Plot of the 150 randomly generated failures along with the median-ranks and the Kaplan-Meier estimates

### Question 1(b): How are the data modeled?

Figures 1 and 2 show that the choice of estimator may not significantly effect the plots or the chosen threshold values. However, the method used to fit the Weibull distribution to the data can have a more pronounced impact on the reliability threshold values derived from the plot. Figure 3 shows the result of fitting the data to a Weibull distribution using least squares (red line) and maximum likelihood (blue line).

A horizontal line is included on plot showing where the 20% failure probability mark intersects both of these lines. The two vertical lines represent the failure times corresponding to a 20% failure probability for the models obtained using least squares (96.38 hours) and maximum likelihood (151.41 hours), respectively. The plot shows a 36.34% difference in the threshold value depending on the modeling approach chosen. **In almost every situation maximum likelihood is a better choice for fitting distributions to reliability data.**

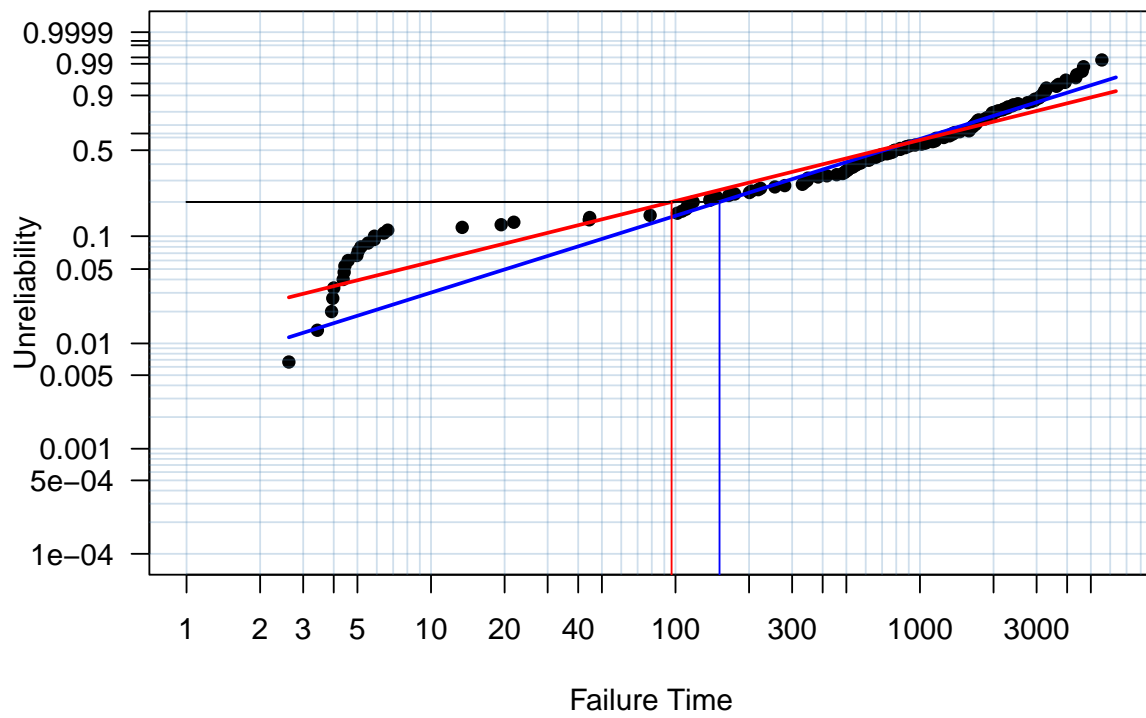


Figure 3: Weibull distribution fits to randomly generated data using least squares estimation and maximum likelihood estimation

## Question 2: How are the failure events defined and how they are included in the analysis?

### Question 2(a) What is meant by “clean failure intervals”?

In previous CBM+ meetings, both PavCon and C3.ai referred to the points shown on Weibull plots as “clean failure intervals”. From a reliability analysis perspective, “failure interval” means that the failure time is only known to have occurred between the start time,  $t_{start}$ , and an end time,  $t_{end}$ . We don’t know the exact failure time, only that the item failed sometime between  $t_{start}$  and  $t_{end}$  and was discovered during an inspection or maintenance action. These failures should be treated as censored observations (aka suspensions) to accurately reflect our uncertainty about when these items failed.

Looking at the points shown in the Weibull charts begs the question: how are these time intervals collapsed down to points? It’s likely that these points represent either the endpoint of the interval (assume that the failure occurred when we discovered it) or the midpoint.

**While it’s not correct to collapse failure intervals down to single points, it’s also not uncommon - particularly when working with messy data. Regardless, the choice of how to represent these interval failures should be documented and applied consistently.**

### Question 2(b) What about the suspensions?

The Weibull plots developed by both PavCon and C3.ai include references the suspensions (or censored observations) in the data. In many cases, the data set contains more suspensions than “clean failure intervals”. **Are these observations included in the analysis?**