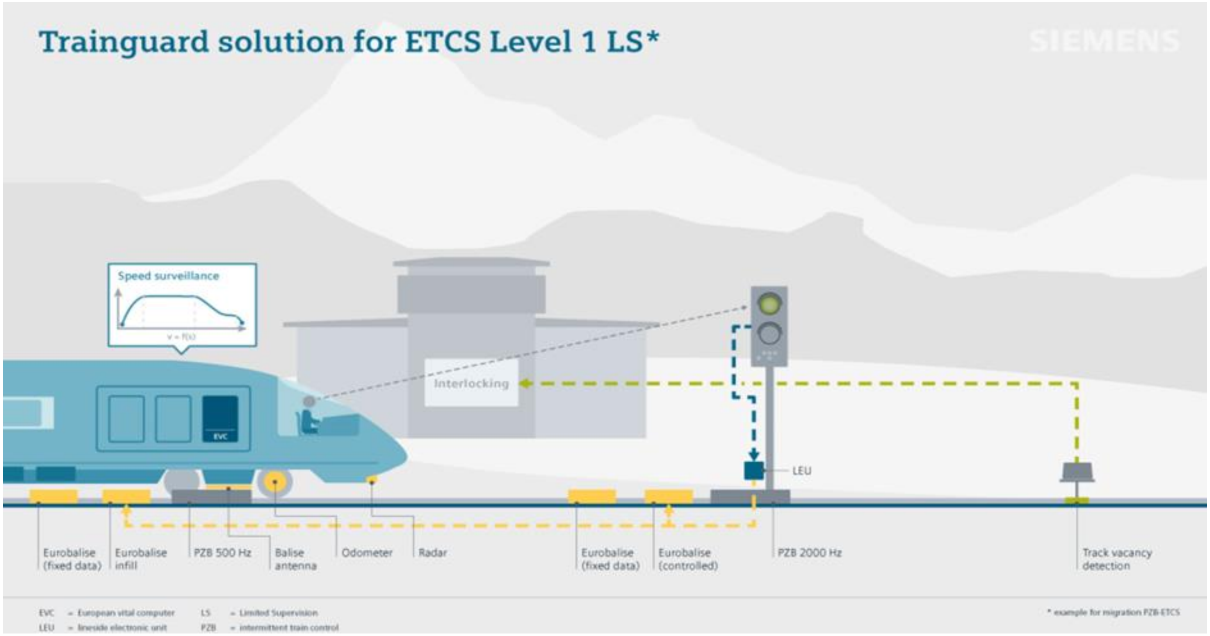
"Equipment Reliability prediction based on lifetime data analysis methodology: The Signaling ETCS Trackside Level 1 Balise case study"



Key Words: Lifetime Data Analysis (LDA), Probability Density Function (PDF), Goodness of fit methods, Maxi likelihood method.

(This paper is fully described in the book "RAMS and LCC engineering: Analysis, Modeling and Optimization (Chapter 4). Author: Dr. Eduardo Calixto)

1 - Introduction

The lifetime data analysis is the basis of reliability prediction as well as other index such as failure rate and unreliability. In order to predict such index for a specific period of time and plot the reliability and failure rate functions, it's necessary to apply different best fit methods in order to know firstly, which probability density function (pdf) fits better with the historical data. Secondly, the PDF parameter definition, which enable the reliability and failure rate function plot and index prediction.

The reliability concept means "probability of one equipment, product or service be successful until a specific time under defined operating conditions. In order to define the equipment reliability is necessary to collect historical failure data.

Therefore, the first step in the lifetime data analysis (LDA) study is to know how failures occur a long time and that's a critical issue for the reliability proper prediction in order to support decisions such as the best time of inspection and preventive maintenance, to check if the equipment is achieved reliability requirement and to supply reliability information to new projects.

To conduct LDA, it is necessary to have historical data at least on the equipment level. Many companies including vendor equipments supplier in the railway industry and also in other

industries do not have an organized database with historical data from their equipment. Therefore, the first step, before the LDA, is to collect the available failure data. The ideal situation is to have a very structured database such as failure report and corrective action (FRACAS). The FRACAS will be presented in another paper but basically, it enables the proper failure historical data report, including the root causes and corrective actions. Such information is the most important source of information to carry out the lifetime data analysis. During the operation phase, the continuous input data in the FRACAS must be carried out. It will enable to perform constantly the LDA for new equipment and component. In addition, it´s also necessary to update the preventive maintenance task as well as work orders.
Nowadays, the solution to put all such information together is the Asset management, which enable not only the FRACAS and maintenance database but also to have operational risk management, asset performance index and monitoring online in only one system. The figure 1 summarizes the LDA application through the asset life cycle.

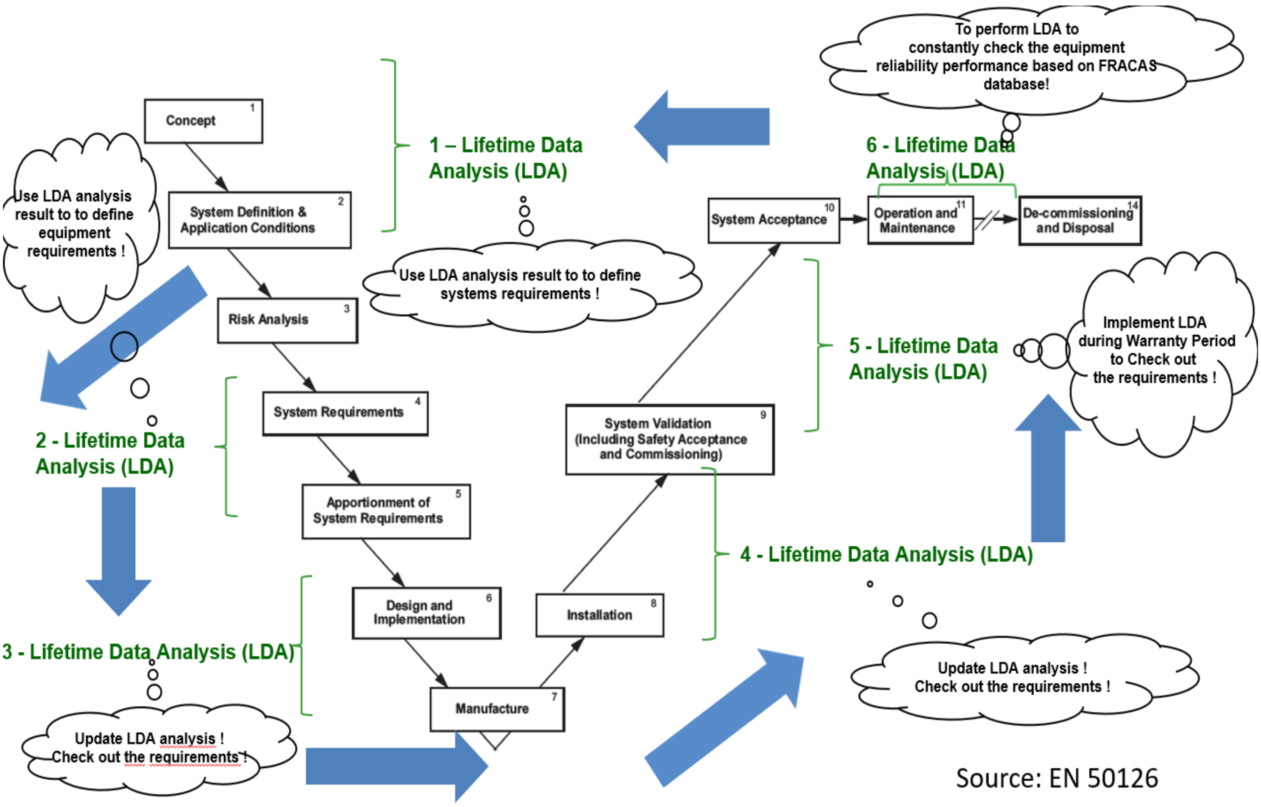


Figure 1 LDA throughout Asset Life cycle

2 – Lifetime data analysis Methodology (LDA)

As discussed before, the LDA aims to predict the equipment and/or component performance index such as reliability, failure rate, based on historical failure data. In order to predict such index, it's necessary to follow the steps shown in figure 2.

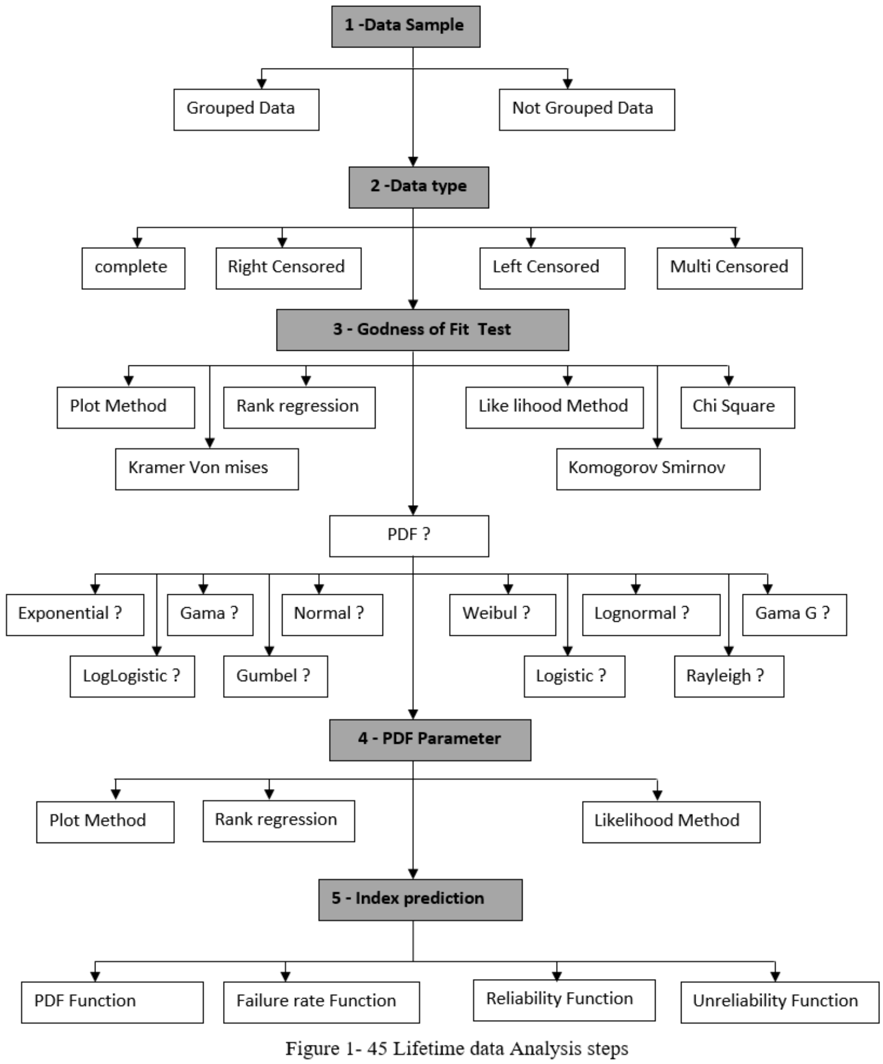


Figure 2 - LDA methodology steps

The first step before starts the LDA methodology is to understand the probability density function (PDF) concept. The PDFs describes graphically the possibility of events occurring over time; in case of the equipment lifetime data analysis, this means, failure or repair time occurrence over time. The Figure 2 shows different shapes of PDFs that represent different types of equipment failure pattern in the oil and gas industry.

The failures may occur at the beginning, during a specific period of time, at the end, or randomly during the equipment life cycle. In some cases, equipment has an expected behavior in terms of failure. The electrical devices, as instance, have expected constant failure rate, but the mechanical component has expected to increase failure rate.

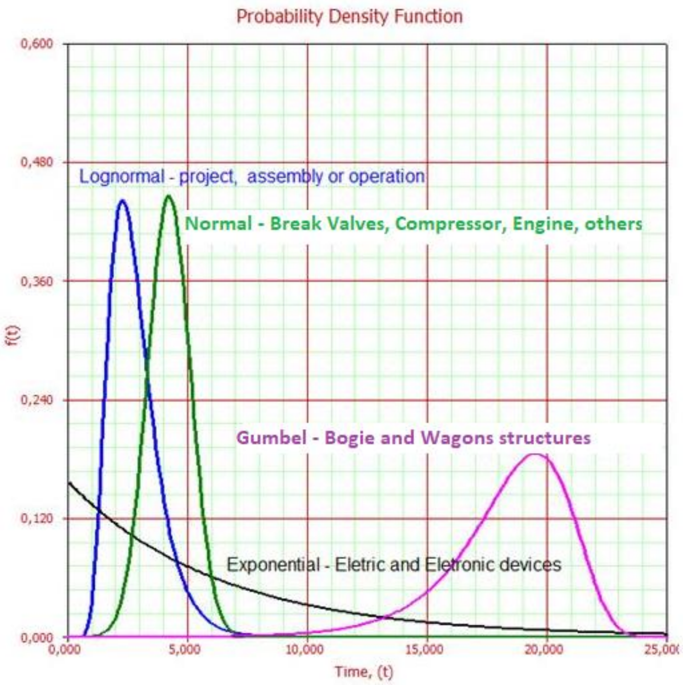


Figure 3 - PDFs and equipment (railway Industry)

In fact, it must be noted that no matter what the PDF shape is, the important issue is to try to understand clearly why the equipment PDF has such shape. It´s also important to validate this information with maintenance professionals and operators who know the equipment issues and troubleshooting. In some cases, some data may be missed or not reported in the historical database files.

Usually, the PDFs for reliability engineering are represented mathematically in most cases as follows:

- Exponential;
- Normal;
- Lognormal;
- Weibull;
- Gumbel.

However, another type of PDF can be applied such as Loglogistic, logistic, Gama, Uniform, Pareto and Rayleigh. The exponential PDF describes random behavior over time and fits well to electrical and electronics equipment best. The normal PDF describes the wear out of some dynamic rotating equipment/ component failures that occurs in specific periods of time with some deviation time. The logistic PDF is similar in shape to the normal PDF but applies a different equation. The lognormal PDF best describes failure that occurs at the beginning of the life cycle that mostly represents failure in a project, startup, installation, or operation. The loglogistic is similar in shape to lognormal but applies a different equation. The Weibull PDF is a generic function and depends on parameter values it assumes the shape of the exponential, lognormal, normal or Gumbel PDFs. The gamma and generalized gamma are also generic

PDFs, which can represent exponential, lognormal, normal, and Gumbel PDFs, depending on parameter value combination. The Gumbel PDFs represents equipment failures that occur at the end of the life cycle, such as corrosion and erosion in a pipeline, vessel, and towers.

Despite being used intensively to describe failure over time, PDFs may also describe repair time, costs, or other variables. For repair time, the lognormal and normal PDFs are most often used by reliability professionals. In case where the lognormal PDF is applied to repair time prediction, it means, most of the repairs are made for short periods of time when performed by experienced employees and take considerable more time when repair is carried out by an inexperienced employee or logistic issues, which cause repair delays. In case where the normal PDF is applied to repair time prediction, it means, the repair is made mostly in a specific period of time with a deviation time.

The PDF shows the behavior of the variable in a time interval, in other words, the chance of such an event occurring in a time interval. So, a PDF is mathematically represented as follows:

$$P(a \le x \le b) = \int_a^b f(x)dx$$

The PDF concept is graphically represented in figure 5, that is the area between interval a and b. However, the cumulative probability of failure is the PDF integration that represents the chance to failure occurs until time t and is represented by the equation below.

$$P(x \le t) = \int_0^t f(x)dx = F(t)$$

The cumulative probability of failure is represented by Figure 6. As discussed before, the reliability is the probability of a piece of equipment, product, or service operating successfully until a specific period of time and is mathematically complementary of cumulative failure probability. Thus, the following equation represents the relation between cumulative failure and reliability (if the two values are added, the result is 100% (or 1). The reliability function is demonstrated in figure 7.

$$R(t) + F(t) = 1$$

$$R(t) = 1 - F(t)$$

$$R(t) = 1 - \int_0^t f(x)dx$$

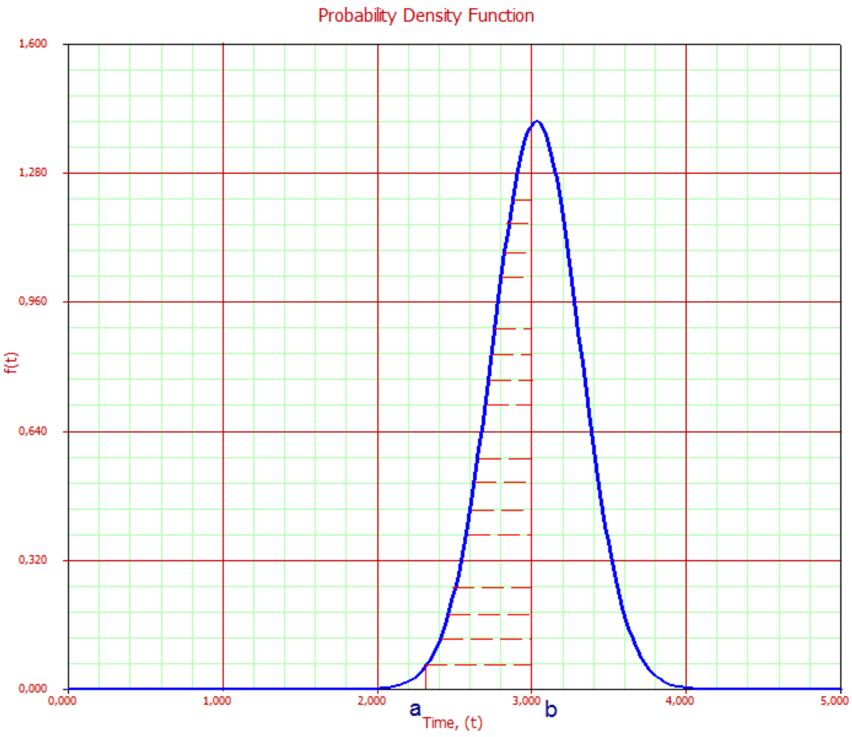


Figure 5 - Probability density function.

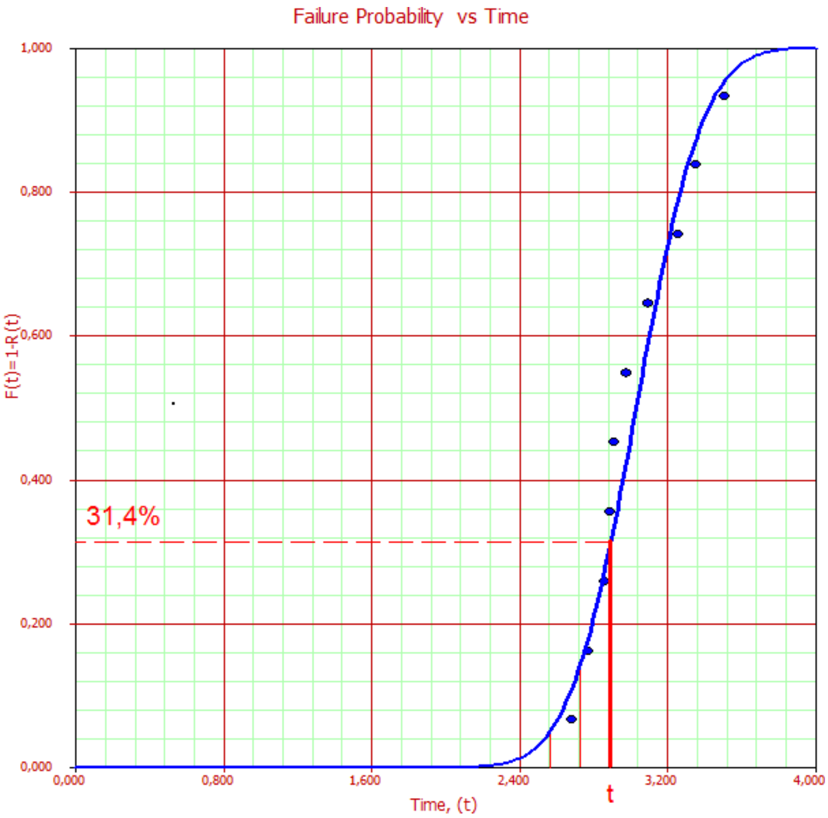


Figure 6 - Probability of failure (from 0 to t = 2.9)

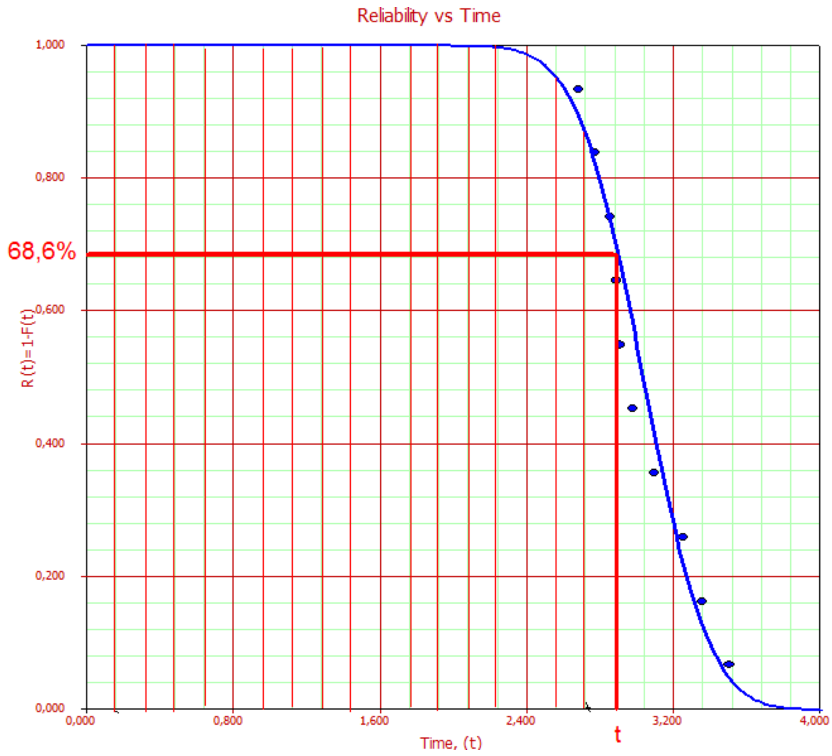


Figure 7 - Reliability function (from 0 to t = 2.9)

Other important index is the failure rate, that is defined by relations between PDF and reliability functions as shows in the equation below.

$$\lambda(t) = \frac{f(t)}{R(t)}$$

Based on the equation above, the failure rate varies along time. To have a constant value, the relation between PDF and reliability must be constant, which happen only in the exponential PDF case. The failure rate function assessment is a very important tool for maintenance and reliability professional cause gives good information how the failure rate behaves along time. The classic failure rate representation is the bathtub curve as shown in Figure 8.

In fact, equipment failure rate is represented for one or two bathtub curve periods. When three periods of equipment life shapes exist, such as the bathtub curve, Weibull 3P (three parameters) is being represented. In Weibull 3P, three pieces of equipment from a common system or three components from one piece of equipment. Thus, the bathtub curve is represented for mixed Weibull, which comprises more than one population; in this case (figure 8), the data of three components are: the early life (β=0,45;η=2,45;γ=0,45) occurs from 0 to 3.8 years, the useful life (β=1,06;η=0,063;γ=0,39) occurs from 3.8 to 7.9 years, and the wear-out (β=49,95;η=8,92;γ=0,14) occurs from 7.9 years on.

Generally lognormal PDF represents well early failures. The exponential PDF represents well random failures. The normal PDF represents well wear out failures. The Weibull 3P may be performing different bathtub curve characteristics. If equipment, component, or product shapes the early life characteristic, in most cases some failure in the project, installation, operation, or startup has happened. If shapes useful life characteristic failures occur randomly and if shapes increasing failure rate that means wear out.

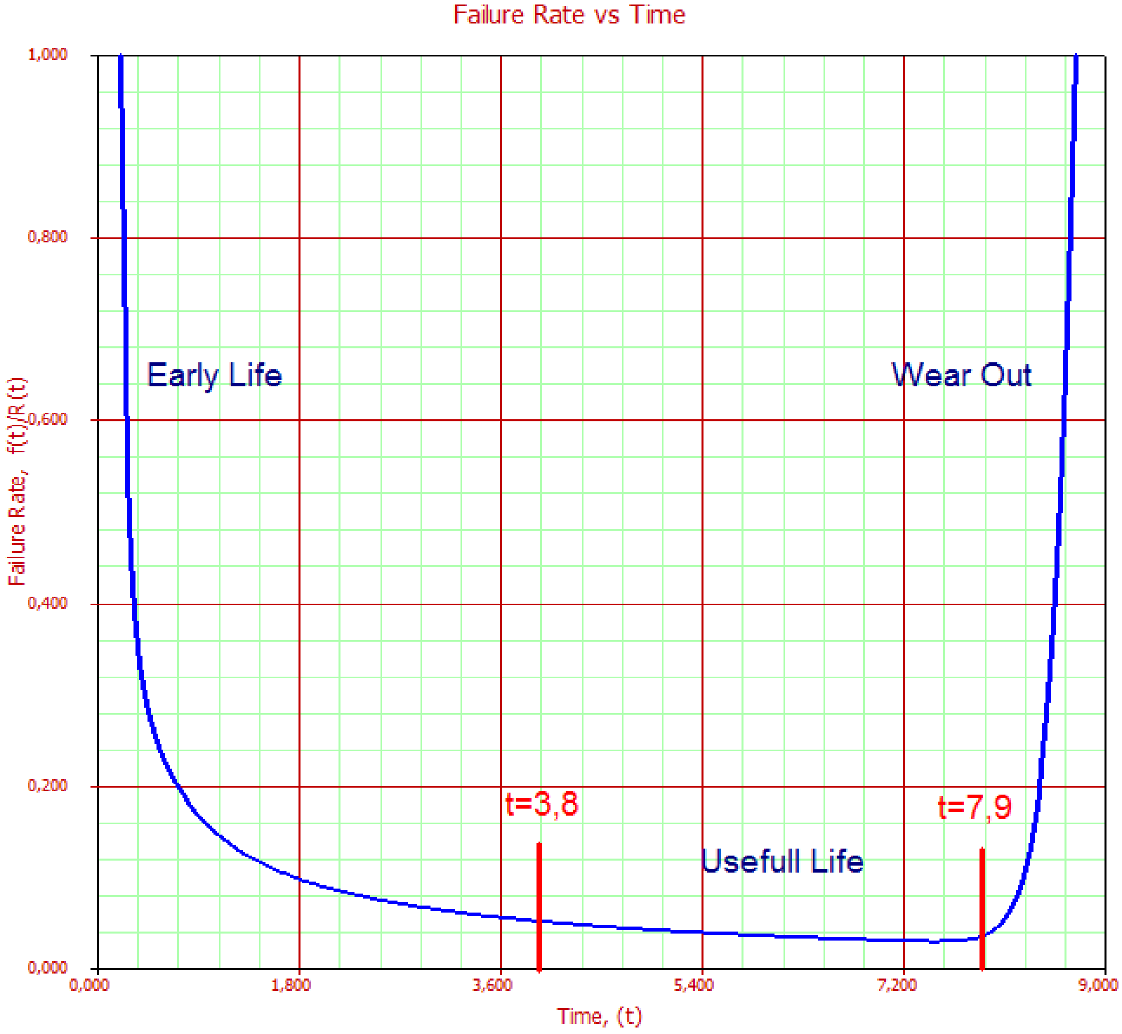


Figure 8 - Bathtub curve.

In fact, one of the most applied PDF during lifetime data analysis is the Weibull function, which can represent exponential, lognormal, or normal shape characteristics. The Weibull PDF can have any of those characteristics, which means a random failure occurrence over the life cycle, or failure occurrence at the beginning of the life cycle with failure time skewed to the right on average with deviation or failure occurrence around a specific period of time centralized in the average with deviation. The Weibull PDF shape behavior depends on the shape parameter (β), which can be:

$0<\beta<1$ (Asymptotic shape)
$\beta=1$ (Exponential asymptotic Shape)
$1<\beta<2$ (Lognormal Shape)
$\beta>2$ (Normal Shape)

Regarding shape parameter, as the beta value gets higher, the PDF shape starts to change from normal shape to Gumbel shape.
The Weibull PDF has three parameters: a shape parameter (β), a characteristic life parameter (η), and a position parameter (γ). If the position parameter is zero, the Weibull PDF has two

parameters. The characteristic life or scale parameter means that 63.2% of failures will occur until the η value, that is, a period of time. The position parameter represents how long equipment has 100% reliability; in other words, there will be no failure until the γ value, which is a certain period of time. In doing so, the Weibull PDF is represented by:

$$f(t) = \frac{\beta}{\eta}\left(\frac{T-\gamma}{\eta}\right)^{\beta-1} e^{-\left(\frac{T-\gamma}{\eta}\right)^{\beta-1}}$$

Where $\beta>0$, $\eta>0$ and $\gamma>0$.

The other important concept in reliability engineering is MTTF, that means the expected time to failure, represented by:

$$MTTF = \int_0^\infty t.f(t)dt$$

In many cases, the MTTF is calculated as an arithmetic average, which is correct only for normal, logistic, or PDFs with such normal characteristics, because in this case mean, mode, and expected time are all the same. Another important concept is the mean time between failure (MTBF) value, which is similar to the MTTF value, but repair time is included in the MTBF case. In many cases in the oil and gas industry, expected time to failure is represented in years and expected time to repair is represented in hours. The MTBF function can be represented as follows:

$$MTBF = MTTF + MTTR$$

$$MTBF = \int_0^T T.f(x)dx + \int_0^t t.f(y)dy$$

where T, is time to failure and t is time to repair. When time to repair is too small compared to time to failure, the MTBF is approximately the MTTF as follows:

$$MTTF >>>> MTTR$$

$$MTBF \approx MTTF$$

$$MTBF \approx \int_0^T T.f(x)dx$$

2.1 – Goodness of fit test

In order to define the PDF parameter as well as the PDF, which best fit on the historical failure data, different Goodness of Fit methods can be applied such Plot Method, Rank regression, Chi square, Komogorov Smirnov, Kramer Von Mises and Maximum Likelihood, which are the most common methods applied.
The maximum likelihood method is one of the most applied methods that fits in all types of data applied to define PDF parameters and verify the historical failure data goodness of fit. Based on the MLE (maximum likelihood estimation) function maximization the PDF parameters are defined. The MLE function is:

$$L(\theta_1, \theta_2, \theta_3 ... \theta_n / x_1, x_2, x_3 ... x_n) = \prod_{i=1}^{n} f(\theta_1, \theta_2, \theta_3 ... \theta_k ; x_i)$$

$$i = 1,2,3 ... n$$

Therefore, the maximum value of the likelihood function related to one parameter is calculated based on the partial derivation of the equation as follows:

$$\frac{\partial(\wedge)}{\partial(\theta_j)} = 0$$

$$j = 1,2,3,4 ... n$$

Where

$$\wedge = \ln L$$

$$L = \prod_{i=1}^{n} f(\theta_1, \theta_2, \theta_3 ... \theta_k ; x_i)$$

$$\ln L = \ln\left(\prod_{i=1}^{n} f(\theta_1, \theta_2, \theta_3 ... \theta_k ; x_i) \right)$$

$$\ln L = \sum_{i=1}^{n} f(\theta_1, \theta_2, \theta_3 ... \theta_k ; x_i)$$

$$\wedge = \sum_{i=1}^{n} f(\theta_1, \theta_2, \theta_3 ... \theta_k ; x_i)$$

To illustrate this method, the balise failure data will be used to estimate the exponential PDF as shows the table 4.3. The balise is the detector which is positioned on the rail to detect the train position and velocity as well as to send train drivers the information about the signaling system. In the exponential PDF case there's only one variable to be estimated, which is l. In doing so, performing the preceding equation steps:

Table 1 Balise database

Equipment	Time to Failure (Years)	Failure Order
Balise 1	0.45	1
Balise 2	1.33	2
Balise 3	2.34	3
Balise 4	3.41	4
Balise 5	3.44	5
Balise 6	4.5	6
Balise 7	4.6	7
Balise 8	4.8	8
Balise 9	4.9	9
Balise 10	5	10

$$\wedge = \ln L$$

$$L = \prod_{i=1}^{n} \lambda e^{-\lambda t} = \lambda^n e^{-\lambda \sum_{i=1}^{n} t_i}$$

$$\ln L = \ln\left(\lambda^n e^{-\lambda \sum_{i=1}^{n} t_i}\right)$$

$$\ln L = n\ln(\lambda) - \lambda \sum_{i=1}^{n} t_i$$

$$\wedge = n\ln(\lambda) - \lambda \sum_{i=1}^{n} t_i$$

$$\frac{\partial(\wedge)}{\partial(\lambda)} = \frac{n}{\lambda} - \sum_{i=1}^{n} t_i$$

$$\frac{\partial(\wedge)}{\partial(\lambda)} = 0$$

$$\frac{n}{\lambda} - \sum_{i=1}^{n} t_i = 0$$

$$\lambda = \frac{n}{\sum_{i=1}^{n} t_i} = \frac{10}{(0.45 + 1.33 + 2.34 + 3.41 + 3.45 + 4.5 + 4.61 + 4.8 + 4.9 + 5)} = 0{,}2876$$

This means 0.2876 failures per year, or MTTF = 3.477 Years.

$$\wedge = n\ln(\lambda) - \lambda\sum_{i=1}^{n} t_i = (10.\ln(0.2876) - (0.2876)(3.477)) = -12.458 - 0.9999 = -13{,}4509$$

The maximum likelihood value represents the balise failure data goodness of fit to the exponential PDF. Whether other PDF is chosen, the higher Maximum Likelihood value shows the one which fits better to the historical data.

3 – Cases Study – Signaling ETCS Trackside: The Balise Lifetime Data Analysis

The next case study concerns the balise, which is part of the ERTMS signaling system (Level 1) as shows figure 4.36. In general terms, the ERTMS encompasses the balises, Lineside Electronic Unit (LEU) and Computer-Based Interlocking (CBI). Basically, the CBI controls the signaling throughout a specific rail region based on information of train speed and location, which comes and back from the train basile antenna, balise and LEU to the CBI. The Balise is the component responsible to send and get the train location and speed. Therefore, in case of failure, such component may trigger some major accident such as collision because the trains are not kept in the safe distance or some wrong signaling information enables some train goes ahead and have a collision with another train. Because of the safety issue, the balise reliability is a critical point in the ERTMS system and must to be accessed since the design phase. Usually, is expected some electric or electronic component failures concern the balises, which is well described by exponential PDF, in other words, random failure. However, any type of PDF may come out from LDA and further investigation based on Root cause analysis is necessary to understanding the cause of failures as well as to verify and validate the reliability performance defined in the warranty contract.

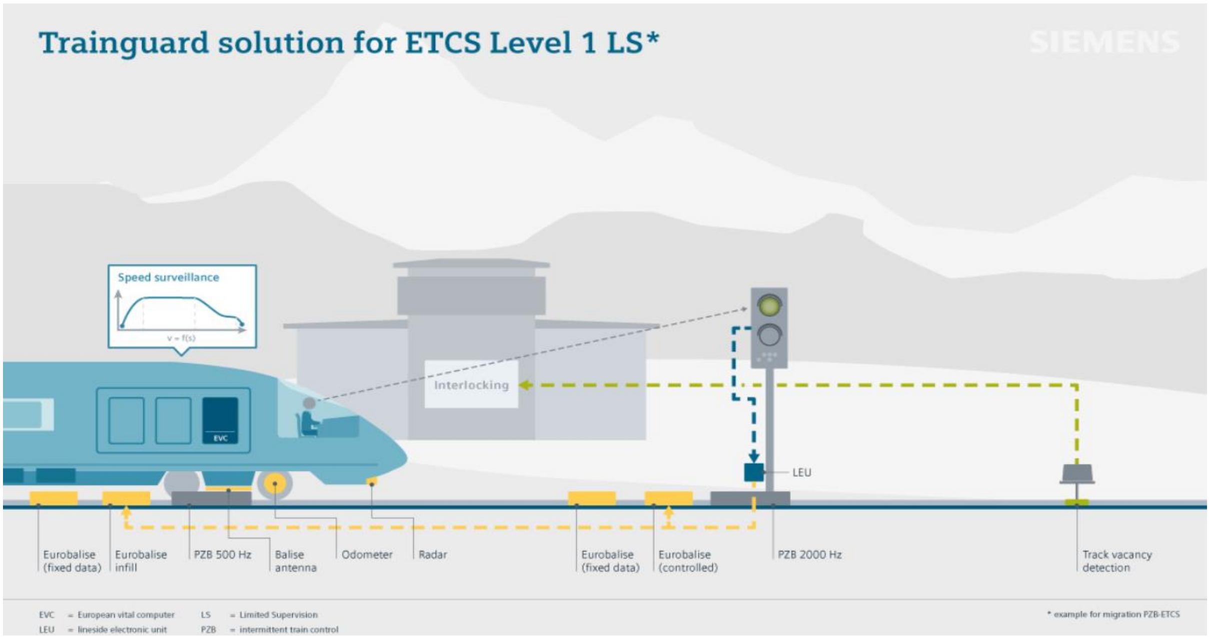


Figure 9 ERTMS level 1. Source: Siemens, 2018

For this case study, let´s consider that the balises are installed in a specific region where specific groups of policies are linked to another group of LEU and at all LEU are Linked to a CBI. For the scope of these LDA case, let´s consider the 70 balises installed in a specific region where the ERTMS is established.
Usually, there are two common mistakes when balises reliability verification and validation is carried out such as:

- To define MTTF as performance index
- To not consider the right censored data, in other words, the basiles which not present failures during the warranty period.

As discussed before, the MTTF may lead a wrong performance assessment. Therefore, the reliability is the performance indexes chosen for this case study. The reliability performance established is, at least, 98% in 2 years.
The other important point, is to take into account all basiles population during LDA, in other words, the failed and right censored ones. In this case study, is being considered the installation of 700 basiles and all will start up its operation at the same time.
In addition, the sample is defined as grouped data, which means, a group of similar basiles will be used to predict the reliability.
The first step is to collect the information from the FRACAS system to check how many basiles failed for 2 years and how many did not fail (right censored). The figure 10 shows the failure historical data in the weibul++ software (reliasoft).

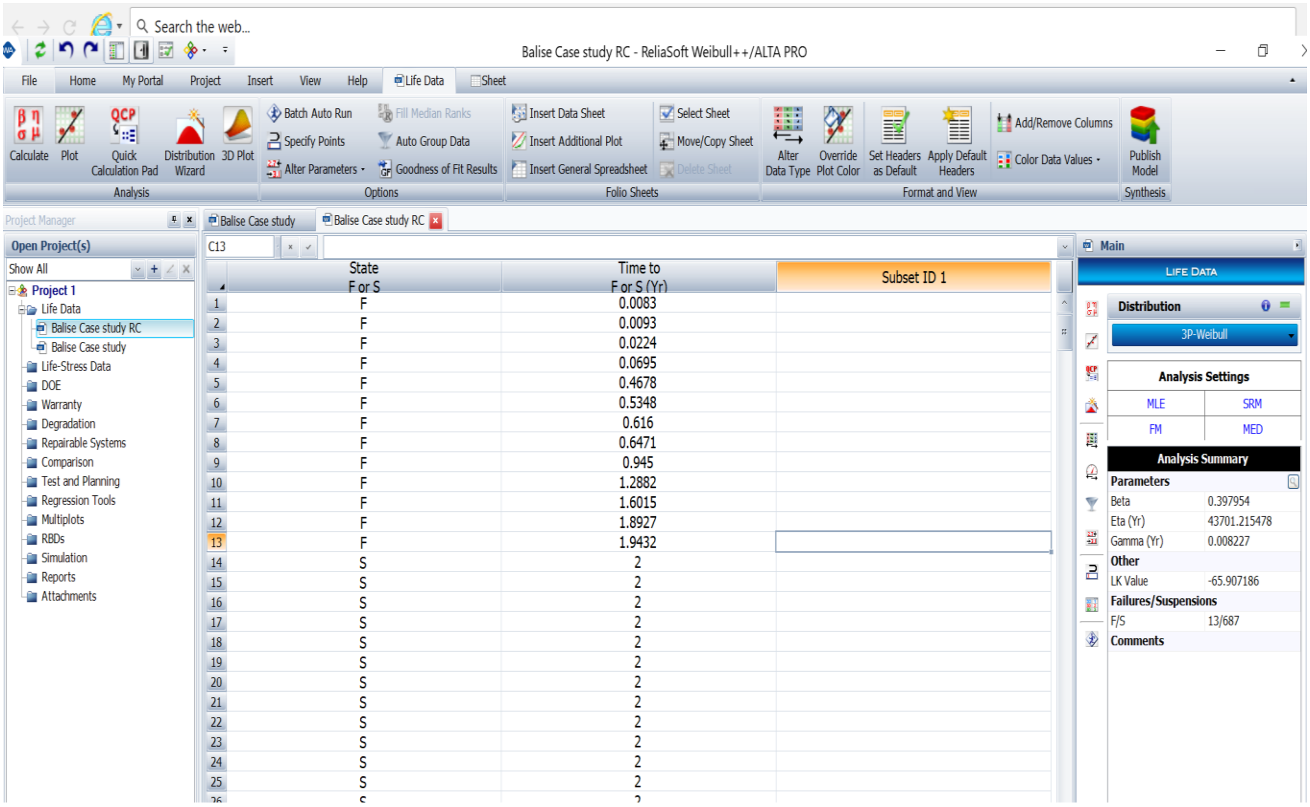


Figure 10 Balises Failures Historical data (2 years operation)

After the performed the MLE test we found out that the Weibull 3P and G-Gama are the most fit functions of the failure data as shows the figure 11. Let´s choose the Weibull 3P to predict the reliability because there´s no any significant difference between the reliability based on these two functions and it´s easier to explain the Weibull 3P based on its parameters results.

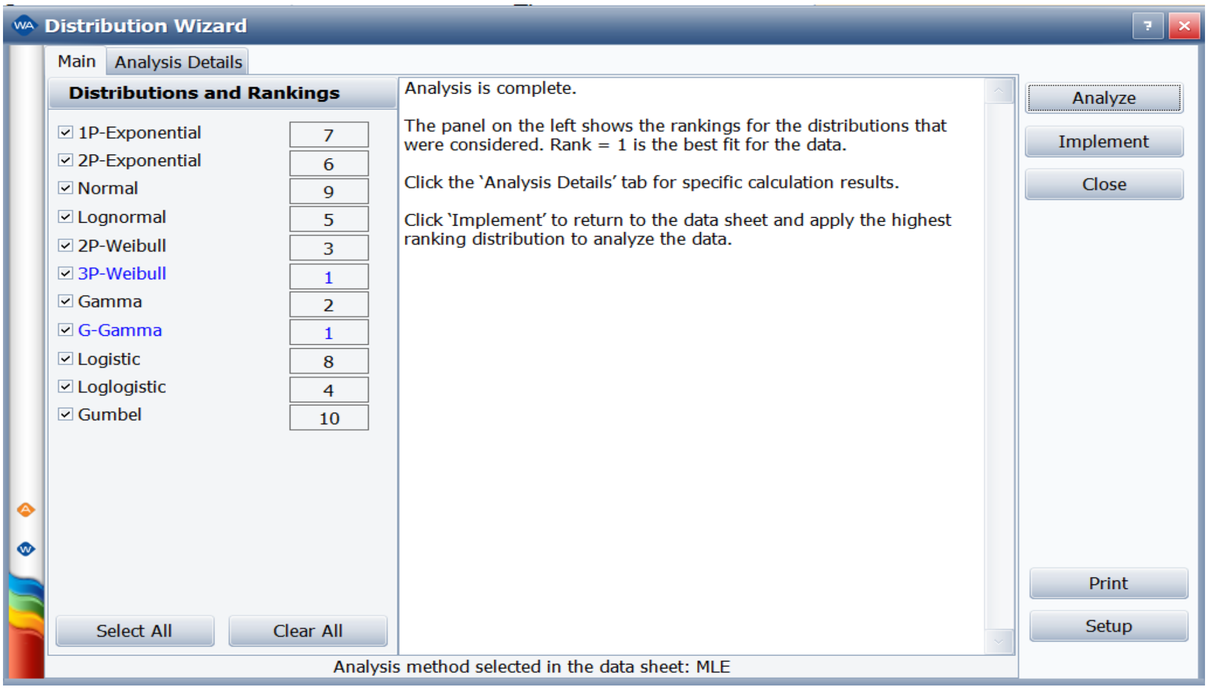


Figure 11 Balises Failures MLE test

The figure 4.39 shows the reliability function based on the Weibull 3P parameter values (β=0.3979, η=43701.21, γ=0.00822). Such parameter values mean:

- Failure occurs during early life period (β=0.3979);
- 63% of failure will occur until 43701 Years;
- The 100% reliability last for a very small period of time, in other words, the reliability decreases soon after operation (γ=0.00822 year).

In addition to using the most fit PDF, the right censored data has a huge impact on the result as shows figure 12. By using the right censored data, the reliability in 2 years is 98.14%, which is higher than 98% in two years reliability, as defined in warranty contract, which validate the reliability performance index.

In case of not considering the right censored data the reliability in 2 years is 7.51%, which doesn't validate the warranty reliability performance target. But in fact, it´s an error of LDA concept application which in many cases is applied in many contracts in railways.

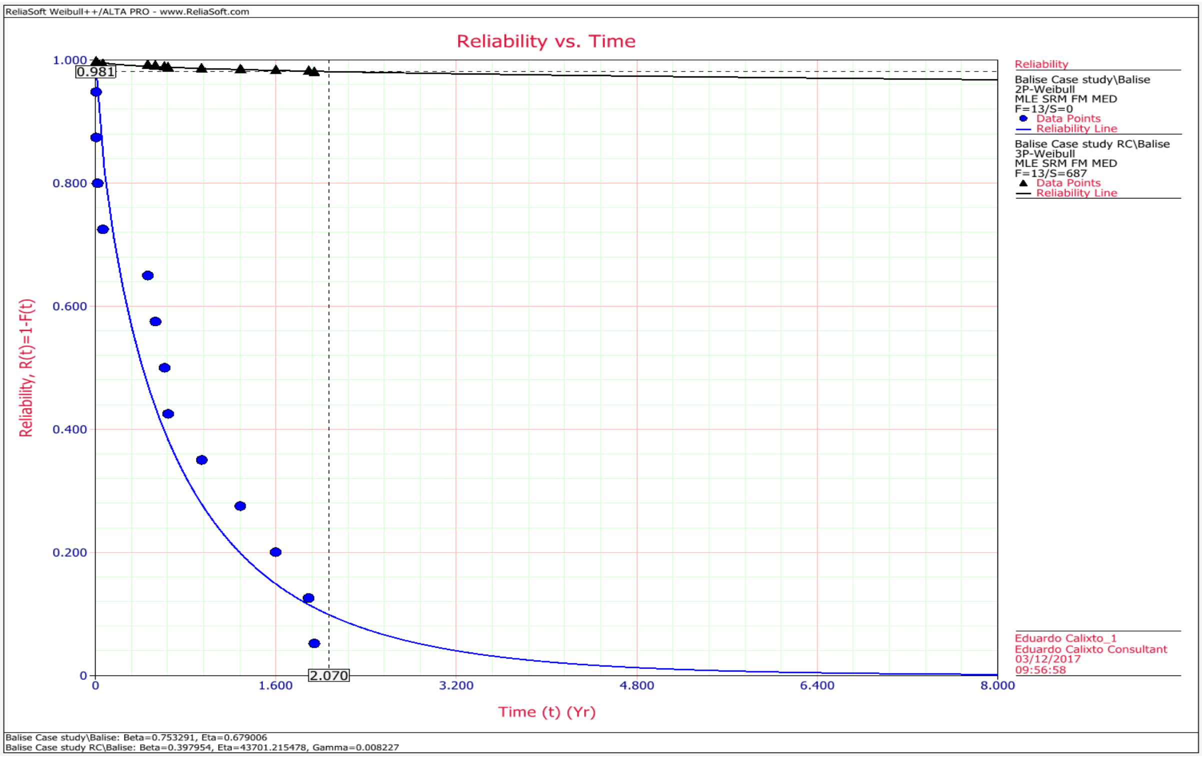


Figure 12 Balises Reliability prediction results comparation

Another error is to consider the MTTF as a performance index based on exponential PDF. If the Exponential PDF is assumed for the balise performance calculation, the MTTF = 106.47 years. By the other hand, if the Weibull 3P is assumed based on MLE test results, the MTTF = 147302.46 years.

The conclusion of these case study is that the assumption of Exponential PDF as well as not considered the right censored data during LDA may lead to a complete wrong result. The consequence is the vendor penalties charge when the performance is achieved. In fact, whenever some equipment fails during the warranty period, it must be replaced, but the penalties must be charged only if the product reliability does not achieve the established target.

4 – Conclusions

The study achieved successfully its objective which was to demonstrate the LDA methodology concept and application. Despite the complexity of the statistic concepts as part of the LDA, the case study demonstrated how easy is the LDA application.

The decision based on the wrong reliability concept can have influence on plant performance as well as operational cost when bad index such as MTTF is the basis for decisions such as inspection and preventive schedule time.

The reliability index may also be applied to compare the different vendor equipment/ component reliability, performance as well as to validate the reliability performance during the warranty period of time. Such very important issues were not the scope of this technical paper but will be discussed in the near future. The detailed LDA approach is described in the book RAMS and LCC engineering: Analysis, Modelling and Optimization including the goodness of fit methods, PDFs (Weibull, Gumbel, exponential, normal and lognormal). Probabilistic Degradation analysis (applied for corrosion and crack prediction) and additional case studies applied to brake, doors and pantograph.