# *“Reliability prediction based on lifetime data analysis methodology: The pump case study”*

***Abstract: The business case aims to demonstrate the lifetime data analysis methodology application from the historical failure data collection to the reliability prediction. Therefore, the concepts of reliability, Probability density function, goodness fit methods, and reliability prediction will be presented as the basis for the LDA application. In order to demonstrate the LDA, the process pump case study will be demonstrated in the end.***

***Key Words: Lifetime Data Analysis (LDA), Probability Density Function (PDF), Goodness of fit methods, Chi-square method.***

***(This paper is full described in the book “Gas and Oil Reliability Engineering: Modeling and Analysis – Chapter 1, case 7.  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 operation 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 in the equipment level.
Many companies including vendors equipment´s supplier in Oil and Gas industry and also in other industries do not have an organized database with historical data for 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 data base such as failure report and corrective action (FRACAS). The FRACAS will be presented in other 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.
However, even if the FRACAS is not implemented, based on equipment files, it´s possible to collect failure historical data. The Figure 1 shows an example of pump file, where different failure mode related to the pumps are pre-defined. In addition, the date where the failure occurred and when it was repaired is also very important to LDA as well as the failure root cause. Such type of reports gives the minimum information necessary to start the LDA and also

is an easier research database for different teams such as inspection, maintenance, process, safety and reliability to understand what happened, why it happened, and to assess the recommendations applied to solve the problems and eliminate the failures.

## Equipment Failure Report

| | | | |
|---|---|---|---|
| Data: 10/12/2004 | Equip Tag: B-114001 A | Managment: Dinamic Maintanance | |
| Ref: R033 | Professional : Alexandre Nunes | | |

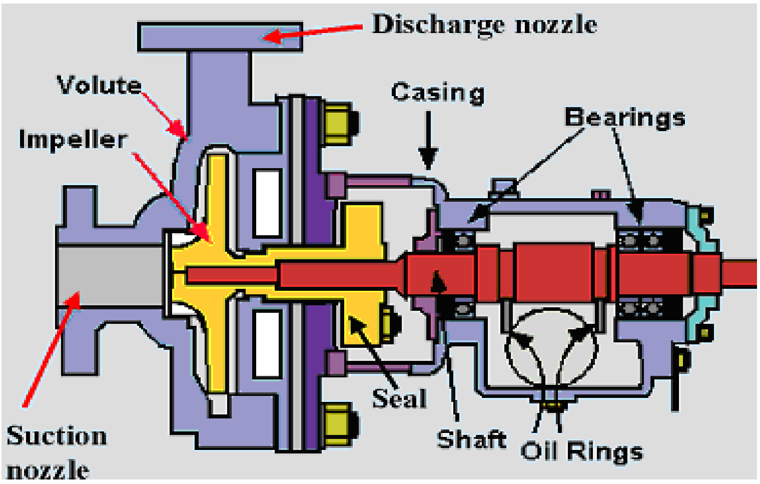
| Type of intervention: | Inspection | Corretive Maintanance | Programmed Maintanance | Predictive Maintanance |
|---|---|---|---|---|
| | | x | | |

| | | | |
|---|---|---|---|
| Data of itervention : | 10/12/2004 | Time of itervention : | 8h |
| Data of start service : | 10/12/2004 | Time of start service : | 9h |
| Data of Finish service : | 12/12/2004 | Time of Finish service : | 10h |

Failure Mode Types

| | Item | Root Cause |
|---|---|---|
| x | 1 - Seal leakage | Pump operation over than specified in procedure |
| | 2 - Bearing | |
| | 3 - Shaft | |
| | 4 - Rotor | |
| | 5 - Eletric Motor | |
| | 6 - Vibration | |
| | 7 - Impeller | |
| | 8 - Rings suction | |
| | 9 - Gaskets | |
| | 10 - Specify other | |

Pump Draw




Signature :

**Figure 1 - Pump file (report page)**

## 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 failure historical data. In order to predict such index, it's necessary to follow the steps shown in figure 2.

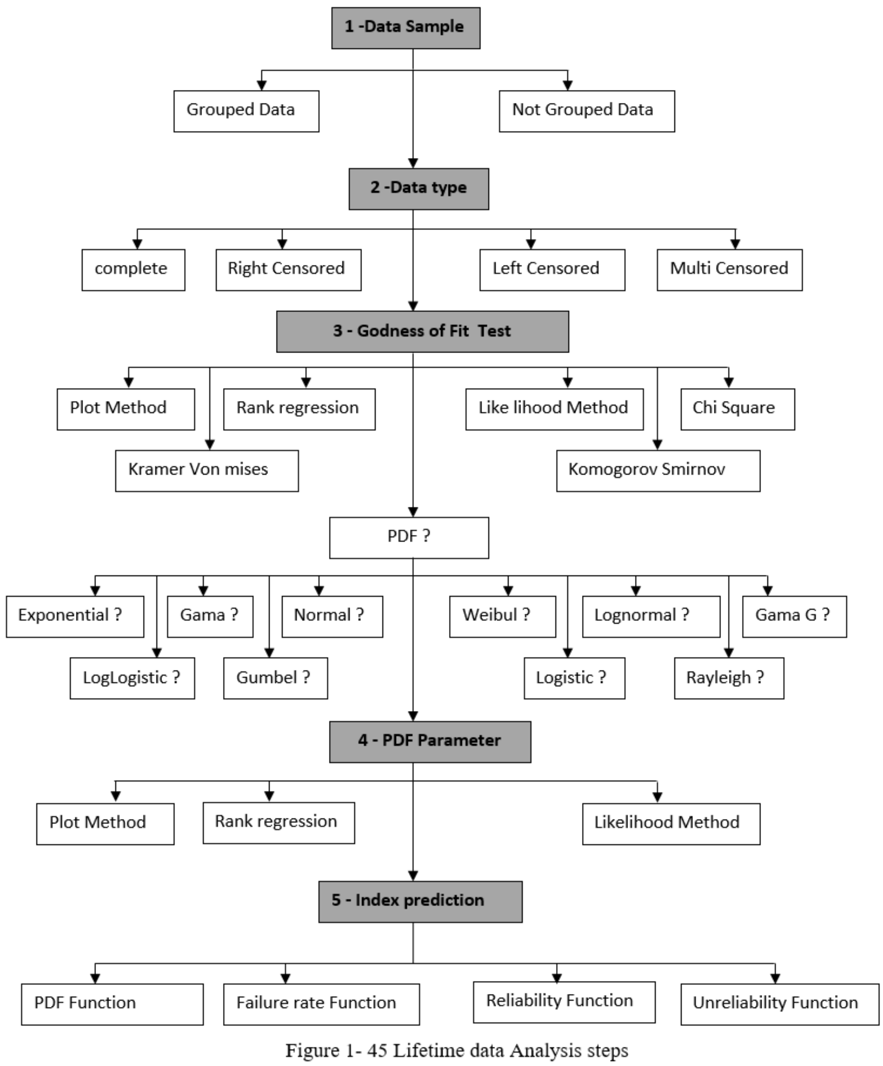



**Figure 2 - LDA methodology steps**

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 mechanical component have expected increasing failure rate.

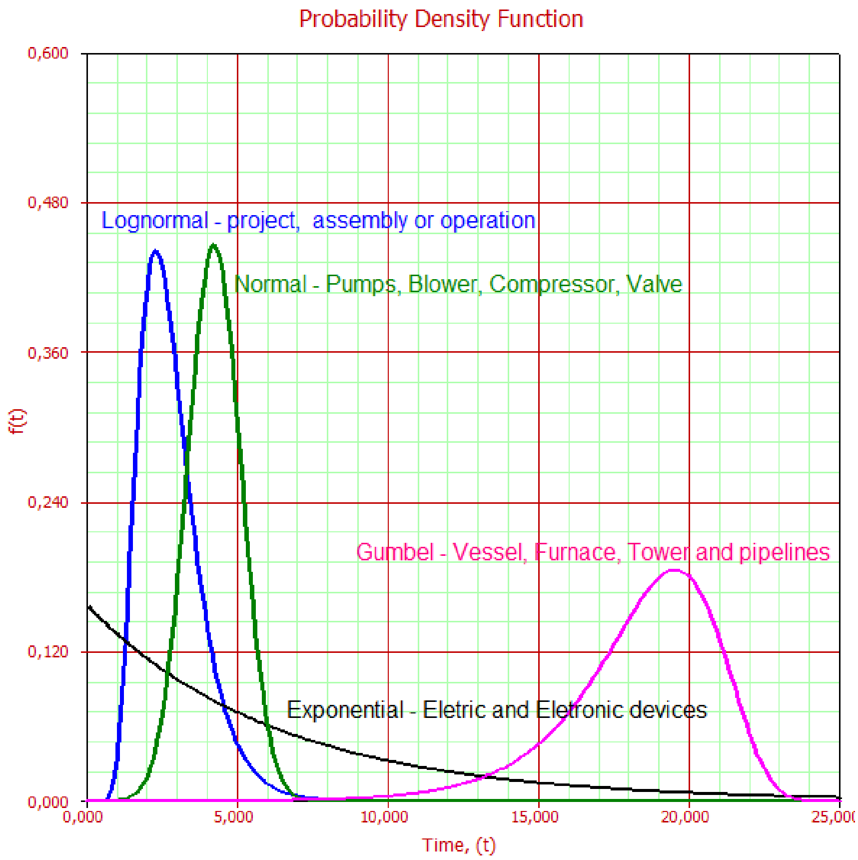



**Figure 3 - PDFs and equipment (Oil and Gas Industry)**

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, other 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 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 different equation. The Weibull PDF is a generic function and depends on parameters values it assumes the shape of 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 values combination. The Gumbel PDFs represent 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, the 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:

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

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.

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

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.

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

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

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

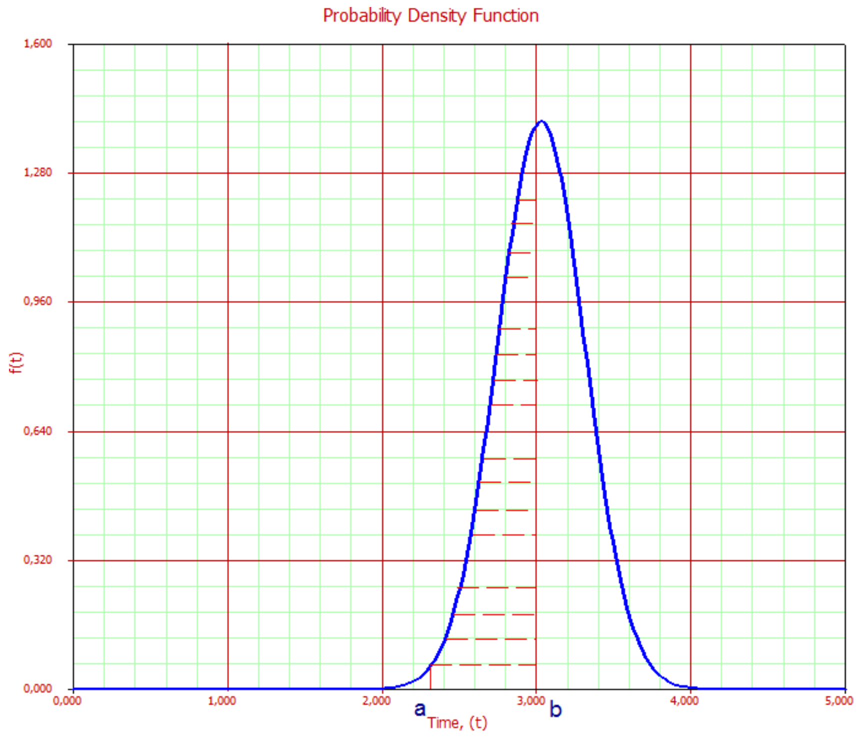



**Figure 5 - Probability density function.**

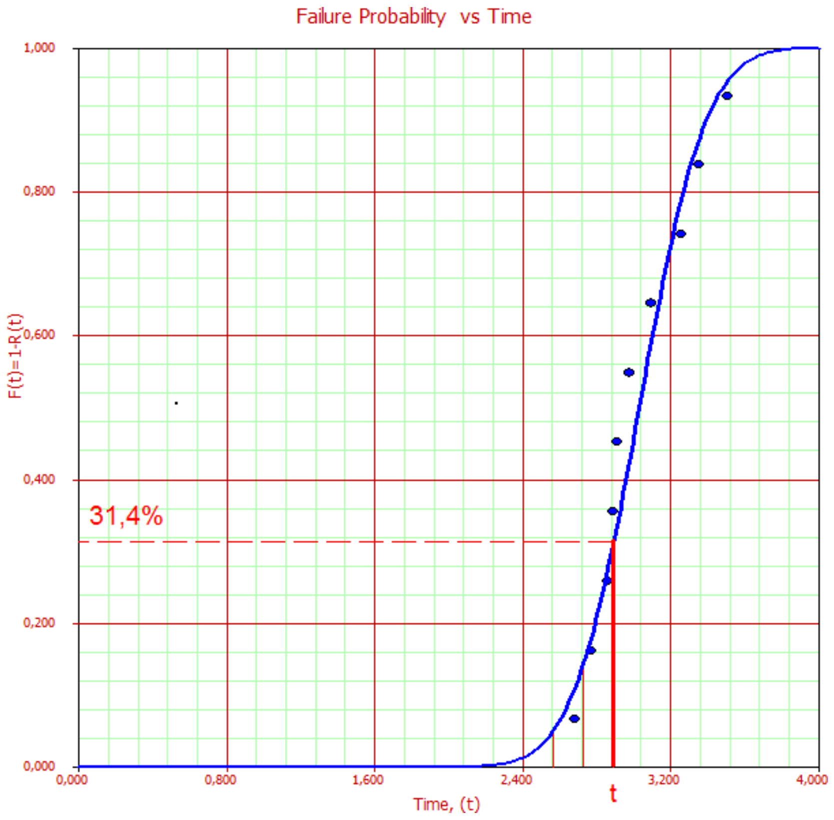



**Figure 6 - Probability of failure (from 0 to t = 2.9)**

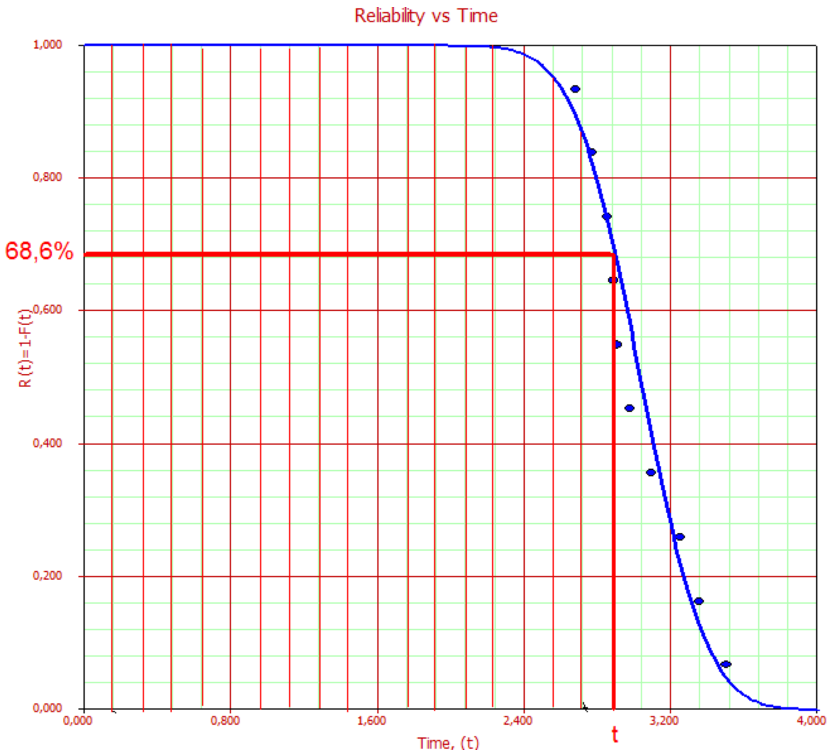



**Figure 7 - Reliability function (from 0 to t = 2.9)**

Other important index is 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 for 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 Weilbull, which comprises more than one population; in this case (figure 8), the data of three components are: the early life ($\beta$=0,45;$\eta$=2,45;$\gamma$=0,45) occurs from 0 to 3.8 years, the useful life ($\beta$=1,06;$\eta$=0,063;$\gamma$=0,39) occurs from 3.8 to 7.9 years, and the wear-out ($\beta$=49,95;$\eta$=8,92;$\gamma$=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.

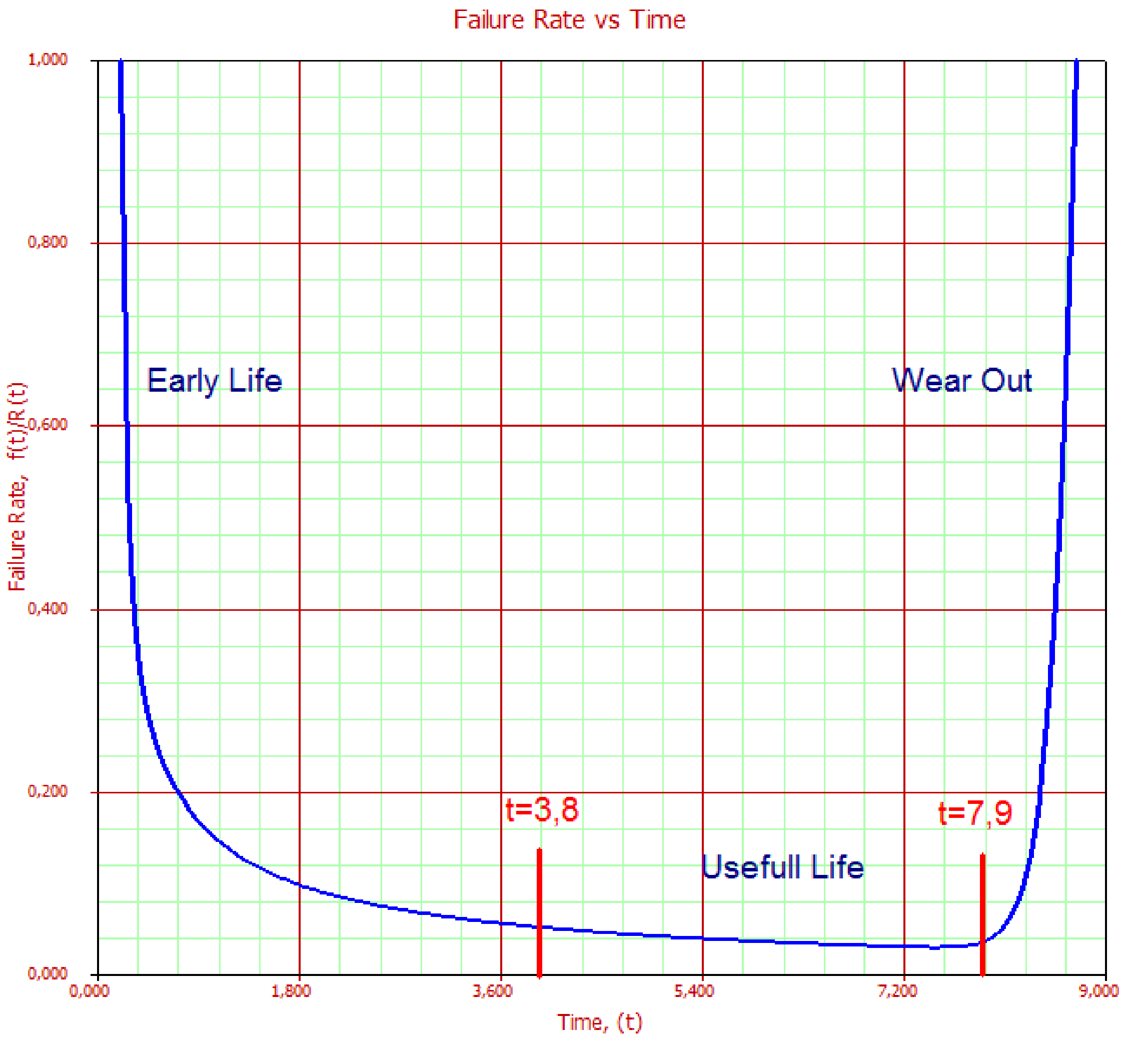



**Figure 8 - Bathtub curve.**

In fact, one of the must 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 β>0, η>0 and γ>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:

$$MTTF >>>> MTTR$$

$$MTBF \approx MTTF$$

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

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 Chi-square method is one of the possibilities to assess the data goodness of fit and find out if such data fits on expected Probability density function (PDF) or not. Such assessment is based on comparing the real data values and predicted values. The parameter used is the frequency, which a measure of goodness of fit account the difference between expected and observed frequencies each squared and divided by the expectation as shows the equation below.

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

Where:

$\chi^2 =$ Chi square value

$O_i =$ Observed frequency

$E_i =$ Expected frequency

The expected value is calculated by equation:

$$E_i = \left( \frac{F(UL_i) - F(LL_i)}{N} \right)$$

F = the cumulative Distribution function for the distribution being tested.

$UL_i$ = the upper limit for class i

$LL_i$ = the lower limit for class i

N = the sample size

The Chi Square value ($\chi^2$) represents the discrepancy between the observed value and the expected value. Thus, the higher is the Chi square value, higher will be a chance to reject the PDF tested.

The second step of the Chi Square is to proceed the hypothesis test based on acceptable error that is defined concerning the risk that is considered acceptable to face in order to take the final decision related to accept or reject the PDF.

The additional parameter is the number of degrees of freedom "α" , usually given by (N – n – 1), where "N" is the number of observations, and "n" is the number of fitted parameters. The

Chi square value related to Degree of freedom (α) must be defined based on the values in the table 1.

**Table 1 - Chi Square critical values**

| α | Chi Square Values | | | | |
|---|---|---|---|---|---|
| | 0.100 | 0.050 | 0.025 | 0.010 | 0.005 |
| 1 | 0.1000 | 0.0500 | 0.0250 | 0.0100 | 0.0050 |
| 2 | 2.7055 | 3.8415 | 5.0239 | 6.6349 | 7.8794 |
| 3 | 4.6052 | 5.9915 | 7.3778 | 9.2103 | 10.5966 |
| 4 | 6.2514 | 7.8147 | 9.3484 | 11.3449 | 12.8382 |
| 5 | 7.7794 | 9.4877 | 11.1433 | 13.2767 | 14.8603 |
| 6 | 9.2364 | 11.0705 | 12.8325 | 15.0863 | 16.7496 |
| 7 | 10.6446 | 12.5916 | 14.4494 | 16.8119 | 18.5476 |
| 8 | 12.0170 | 14.0671 | 16.0128 | 18.4753 | 20.2777 |
| 9 | 13.3616 | 15.5073 | 17.5345 | 20.0902 | 21.9550 |
| 10 | 14.6837 | 16.9190 | 19.0228 | 21.6660 | 23.5894 |

Based on the table 1, the value of $\chi^2 = 0.1$, considering α=1 and 90% of confidence level. The degree of freedom (α) depends on the number of observations and fitted characteristic as discussed above.
The last step is to apply the hypothesis null test (H0) to decide if the data (PDF) are accepted or rejected. Therefore, if:

$$H_0 : \chi^2 < \chi_\alpha{}^2$$

The hypothesis is accepted, and the PDF has a goodness of fit to the historical data collected.
In order to exemplify the Lifetime data analysis methodology a case study will be carried out in the software CAFDE from BQR.

### *3 – Cases Study - Pump Lifetime Data Analysis*

The case study describes the real case of proper reliability application concept. The LDA of the pump was requested for the Maintenance Manager because he cannot explain why they are not able to avoid the pump failure based on their schedule inspection and preventive maintenance. Based on maintenance engineer calculation, the inspection and preventive maintenance have to be carry out soon after the 19000 hours, which is defined based on the MTTF calculation (MTTF=20000 hours).
Therefore, the reliability engineering group carry out the lifetime data analysis applied to the Propylene Plant´s Pump (the feed pump), concerning the component failures described in the historical database. Because in refinery plants each process plant has different operational condition, product, configuration and design set up such as pressure, temperature and flow, the pump is considered a not grouped sample.
Based on failure historical database, the data are complete, which means, all failure available has a defined date and the pump failed during the period assessed. The table 2 shows the failure

historical database from two main pump´s component such as seal and bearing. The pump reliability is highly affected by these two components. It´s also important to consider the interval of time, which such component were replaced based on preventive maintenance. Therefore, the additional important information on the database is about maintenance, cause of failure, the consequence and th proposed solution. In fact, in order to perform a reliable lifetime data analysis, it´s necessary to have a clear and reliable database as well as such additional information to prevent failures in the future.

**Table 2 – Pump Failure historical data**

| Equipment | Component | Failure | Repair Start | Repair finish | Maintenance type | Cause | Consequence | Solution |
|---|---|---|---|---|---|---|---|---|
| B-31005 B | bearing | 25/02/1994 | 26/02/1994 | 03/03/1994 | Corretive | Wear out | High Vibration | Replace |
| B-31005 B | bearing | 09/11/1998 | 09/11/1998 | 17/11/1998 | Corretive | Wear out | High Vibration | Replace |
| B-31005 B | bearing | 21/12/1998 | 22/12/1998 | 22/12/1998 | Predictive | Wrong design | High Vibration | Redesign |
| B-31005 B | bearing | 02/03/2001 | 02/03/2001 | 14/03/2002 | Corretive | Installation error | High Vibration | Replace |
| B-31005 B | bearing | 04/09/2002 | 06/09/2002 | 16/09/2002 | Predictive | Wrong design | Loss of performance | Redesign |
| B-31005 B | Seal | 07/01/2003 | 08/01/2003 | 17/01/2003 | Corretive | Wear out | Leakage | Replace |
| B-31005 B | Seal | 25/01/2003 | 25/01/2003 | 27/01/2003 | Corretive | Installation error | Leakage | Replace |
| B-31005 B | Seal | 17/06/2003 | 18/06/2003 | 20/06/2003 | Corretive | Operation error | Leakage | Replace |
| B-31005 B | Seal | 25/07/2003 | 25/07/2003 | 26/07/2003 | Corretive | Installation error | Leakage | Replace |
| B-31005 B | Seal | 23/08/2003 | 24/08/2003 | 29/09/2003 | Corretive | Installation error | Leakage | Replace |
| B-31005 B | bearing | 25/10/2004 | 26/10/2004 | 28/11/2004 | Corretive | Wear out | Loss of performance | Replace |
| B-31005 B | Seal | 10/01/2005 | 10/01/2005 | 11/01/2005 | Corretive | Wear out | Leakage | Replace |

The next step is to perform the goodness of fit test and in most of the cases, a reliability software is the most indicated solution to support such analysis in order to have a faster and more reliable result.

Therefore, the CAFDE software from BQR Reliability Engineering Ltd was applied. The CAFDE Software has the Chi-Square methods to perform the Goodness of fit test and the likelihood method to predict the PDF parameters. Before performing such test, it´s necessary to organize the information collected on failure historical data base on excel format in order to export directly to the CAFDE. The table 3 shows the data organized on excel format to import directly to CAFDE software. The further step is to perform the Goodness of fit test and predict the PDF parameter as shown in figure 9.

The figure 9 shows the possible type of PDFs such as Normal, Lognomal, Exponential, Weibull, Uniform, Pareto and Rayleigh.

**Table 3 - Pump Failure historical data – CAFDE template**

| Data Start Date | 01/01/1990 | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Off Date | Off Type | Off description | System Serial Number | System Location | Maintenance Start Date | Maintenance Finish Date | On Date | LRU Name | LRU Serial Number | SRU Name | SRU Serial Number | Component Name |
| 25/02/1994 | failure | description 1 | A3 | 1401 | 25/02/1994 | 03/03/1994 | 03/03/1994 | bearing | B-31005 B | bearing | 2 | bearing |
| 09/11/1998 | failure | description 2 | A3 | 1401 | 09/11/1998 | 17/11/1998 | 17/11/1998 | bearing | B-31005 B | bearing | 2 | bearing |
| 21/12/1998 | failure | description 3 | A3 | 1401 | 21/12/1998 | 22/12/1998 | 22/12/1998 | bearing | B-31005 B | bearing | 2 | bearing |
| 02/03/2001 | failure | description 4 | A3 | 1401 | 02/03/2001 | 14/03/2002 | 14/03/2002 | bearing | B-31005 B | bearing | 2 | bearing |
| 04/09/2002 | failure | description 5 | A3 | 1401 | 04/09/2002 | 16/09/2002 | 16/09/2002 | bearing | B-31005 B | bearing | 2 | bearing |
| 07/01/2003 | failure | description 6 | A3 | 1401 | 07/01/2003 | 17/01/2003 | 17/01/2003 | Seal | B-31005 B | Seal | 1 | Seal |
| 25/01/2003 | failure | description 7 | A3 | 1401 | 25/01/2003 | 27/01/2003 | 27/01/2003 | Seal | B-31005 B | Seal | 1 | Seal |
| 17/06/2003 | failure | description 8 | A3 | 1401 | 17/06/2003 | 20/06/2003 | 20/06/2003 | Seal | B-31005 B | Seal | 1 | Seal |
| 25/07/2003 | failure | description 9 | A3 | 1401 | 25/07/2003 | 26/07/2003 | 26/07/2003 | Seal | B-31005 B | Seal | 1 | Seal |
| 23/08/2003 | failure | description 10 | A3 | 1401 | 23/08/2003 | 29/09/2003 | 29/09/2003 | Seal | B-31005 B | Seal | 1 | Seal |
| 25/10/2004 | failure | description 11 | A3 | 1401 | 25/10/2004 | 28/11/2004 | 28/11/2004 | bearing | B-31005 B | bearing | 2 | bearing |
| 10/01/2005 | failure | description 12 | A3 | 1401 | 10/01/2005 | 11/01/2005 | 11/01/2005 | Seal | B-31005 B | Seal | 1 | Seal |

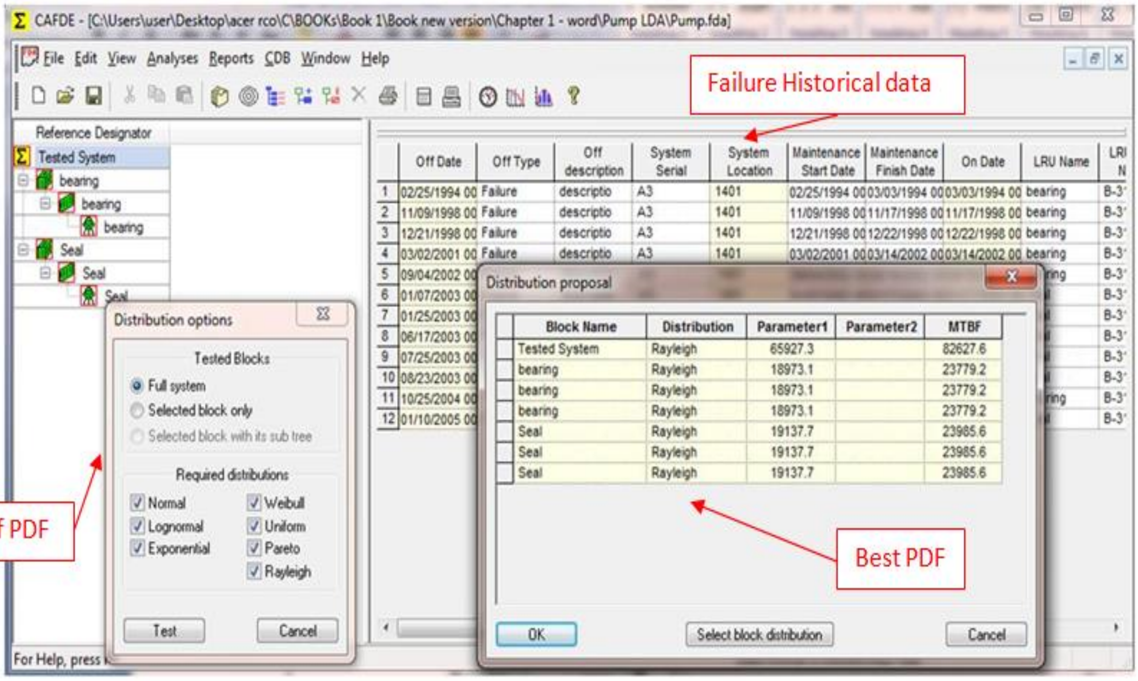



**Figure 9 - Pump Failure historical data – CAFDE LDA**

The figure 10 and 11 shows the final PDF parameter estimation for the two most critical components such as bearing and seal respectively. On the left of figures 10 and 11 is demonstrated the level of significance of Chi Square test for each type of PDF. On the bottom of figures 10 and 11 are the parameter α values.

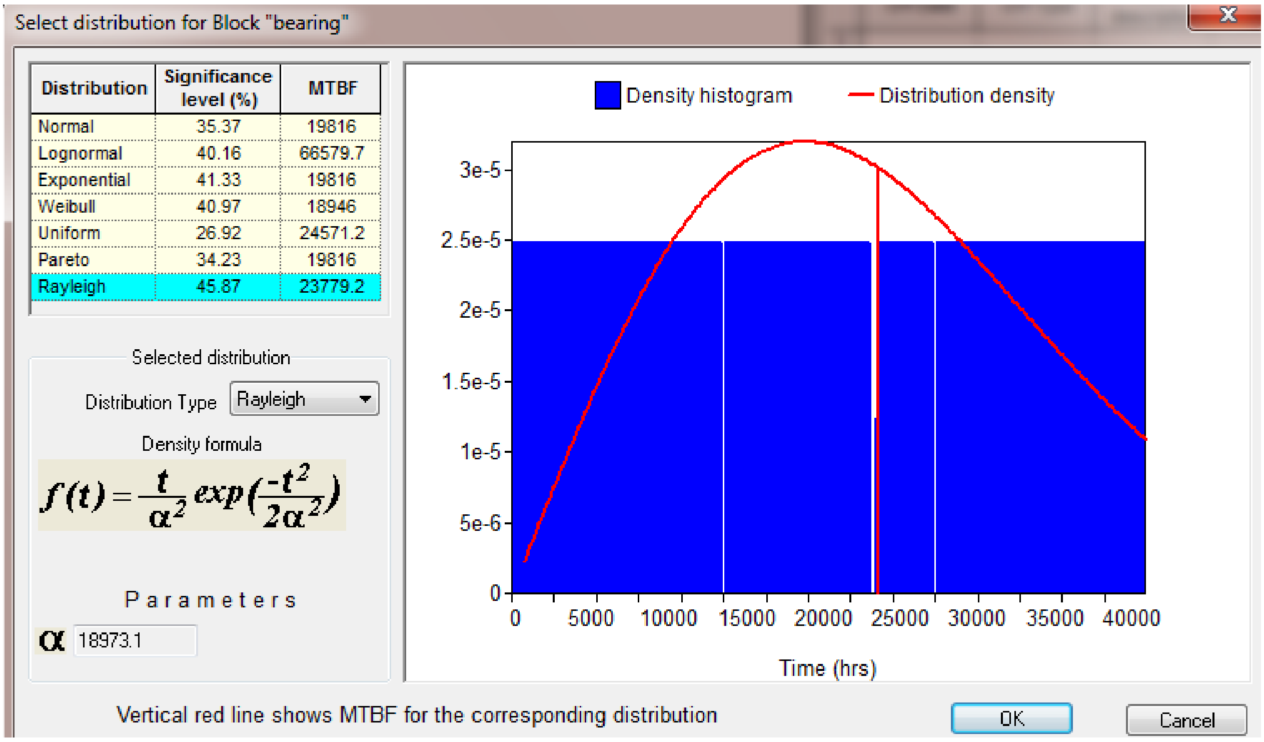



**Figure 10 – Pump Bearing LDA – CAFDE template**

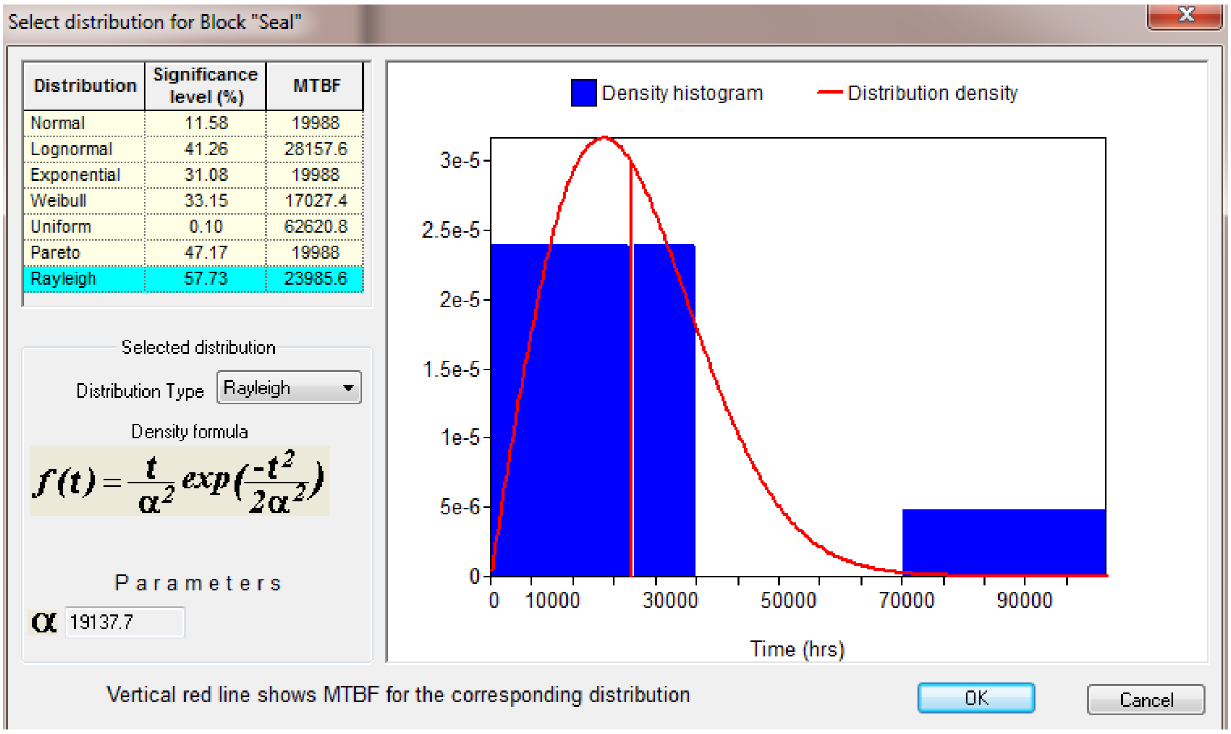



**Figure 11 - Pump Seal LDA – CAFDE template**

After the PDF parameter estimation, the final step is to predict the reliability and failure rate function as shown in the figures 12 and 13 respectively for the bearing component.

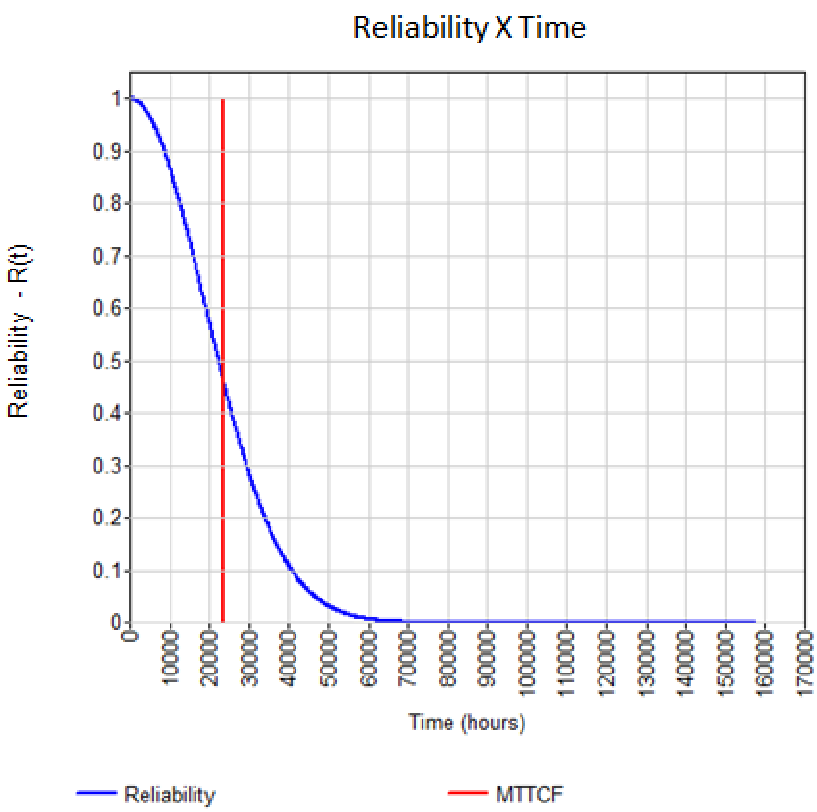



**Figure 12 - Bearing Reliability function**

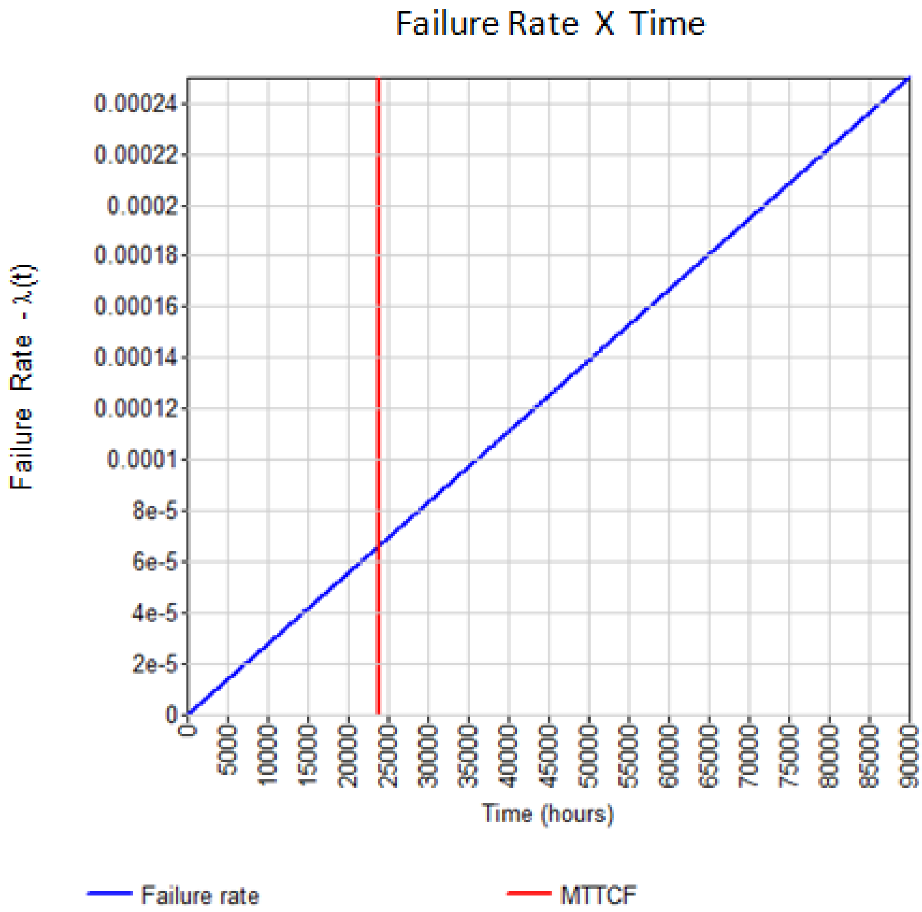



**Figure 13 – Bearing Failure Rate function**

The Lifetime data analysis indicates the low reliability in 20000 hours (R(20000)=49%), which is a too late to perform some preventive maintenance and try to recover the pump reliability to the state as good as new (or close of such state).
Such result contradicts the preliminary definition of the maintenance group that defined to implement preventive maintenance in such pump soon after 20000 hours based on MTTF calculation.
In fact, if the maintenance group keeps with such decision, they will not be able to avoid the pump failures and will keep the operation cost, higher than necessary for the pump corrective maintenance is more expensive than the preventive maintenance and have also a higher downtime, which expose the whole plant to a risk of shutdown for a longer.
Based on such decision the Inspection and preventive maintenance were redefined for a period of every 8760 hours, which allowed to avoid the bearing as well as to keep the pump under as good as new state. Despite a good preventive maintenance policy based on LDA, the managers are not happy with the bearing performance, which suppose to operate at least 17520 hours without failure. Therefore, a new LDA study will be implemented to define the best bearing reliability performance among the pump suppliers.

## *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 was not the scope of this technical paper but will be discussed in the near future.