The Concept of OREDA on Data Treatment for Equipment

Renato Fernando Mendes, PETROBRAS/CENPES, Sr. Consultant

ABSTRACT

Safety, Maintenance and Operations are all connected since without well maintained and operated equipment an industry will not get safe/reliable performance on their critical ones. To get common metrics for critical equipment, during early 80s, some Oil & Gas companies got together for a continuous JIP-Joint Industry Project named as OREDA – Offshore and Onshore Reliability Data. The main purpose of this organization project was to establish a comprehensive databank with reliability and maintenance data for exploration and production equipment from wide variety of installations, equipment types and diverse operating conditions. Offshore topside and subsea equipment units were primarily covered and some onshore were secondly considered.

It is introduced the format of OREDA based on a structured database for benchmarking process to O&G companies, as well as been reference to create the bases for ISO 14224, 1st Edition, on 1999. A more internal and formal part for reliability estimation of equipment units are presented for a broad understanding of its concept and applicability for equipment benchmark and system RAMS. For robust view of this application, some examples were developed.

1. INTRODUCTION

The OREDA (Offshore and Onshore Reliability Data) database is an industry project effort between several Oil Companies such as: BP, ENGIE, ENI, Gassco, Petrobras, Statoil, Shell, Total to collect and analyze field reliability data for offshore and onshore equipment. The project has been carried out in twelve phases beginning in 1981 up to today.

The main goal of the OREDA project is to improve safety and cost-effectiveness in design and operations of Oil & Gas industry through collection and analysis of maintenance and operational data, establishment of a high-quality reliability database, and exchange of reliability, availability, maintenance and safety (RAMS) technology among the participating companies. OREDA has published six editions of its reliability data handbook in 1984, 1992, 1997, 2002, 2009, and 2015.

All collected data are stored in a database containing data from 278 installations, 17.000 equipment units with 39.000 failure and 73.000 maintenance records. The databank also includes subsea fields with more than 2.000 years of operating experience. DNV GL has been the Project Manager since 2009, and before of that SINTEF was conducting this Joint Industry Project.

The project is now closing its 12th phase (2015-2017). In this phase, data is collected by more automatic means, and the data collection specification, and the data collection and analysis software program have been revised on 11th phase. Petrobras had jointed OREDA at 10th and 11th phases and now is continuing within the project.

The participating oil companies use the data on the development of new oil fields and improving existing facility operation. The reliability data are typically used as input for safety and reliability analyses. Some benefits are: safer operations, increased production availability, and optimised maintenance. Analysis of reliability data is one of the key factors in choosing cost-effective solutions. Two examples recently reported, have shown that savings of about $60 millions can be achieved compared with the cost of the original design. Lately experiences and data from this project are also exchanged with the manufacturers in order for them to improve future designs.

2. DATA CONFIDENTIALITY

On project OREDA efforts were considered to design a secure framework to guarantee an anonymous database for operators, and been on acceptable level of confidentiality for the joined companies. Only generic data will be handed over by the main contractor DNV GL for merging of data from the different participants and eventually input to the handbook. The main contract has responsibility to conduct the OREDA phases doing administration, developments and data quality production for specific work phases.

For data transfer to OREDA, each participating company has the choice to get themselves data from their databank or to leave a subcontractor responsible for the data collection and processing within the company.

On the other hand, full equipment description is possible to be found on the database and this is an opportunity to develop a feedback way to equipment manufactures. For this purpose, there is annual meeting with subsea vendors and OREDA staff during which, among other things, main performance comparisons are done about their equipment units.

3. OREDA DATABASE

The OREDA records anonymously data per operators and installations; each individual item (e.g. a gas turbine) occupies a single inventory record in the database. This record contains a technical description (e.g. manufacturer information) plus operating and environmental conditions. For each inventory, all failure events are stored. Each failure event is identified by item name, date of failure, failure impact, failure mode, failure cause etc. The maintenance records contain data on corrective maintenance linked to the corresponding failure record, and data on preventive maintenance linked to the corresponding inventory record.

The OREDA experience on collecting reliability and maintainability data allowed the first Edition of ISO 14224 “Collection and exchange of reliability and maintenance data for equipment” in 1999. The standard is applicable to all equipment types used in the petroleum and natural gas industry, such as process equipment (used on onshore and offshore installations), subsea equipment, well-completion equipment and drilling equipment.

4. DATA TREATMENT

In order to estimate equipment reliability and maintainability, the project OREDA gives advantage to partners to share their data and reach large data sets for similar equipments. Having these data, there is a need to define methods for data treatment.

When failure rates are estimated by OREDA, the following approach and assumptions are made:

- It is done an inventory filter, where inventory units are assumed to have similar taxonomy and operation even if there are not from same installation or operator;
- Data failure from similar equipment units inside same installation are treated as Homogeneous Poisson Process (HPP) and it is the same as to say that failure time series are exponentially distributed with the same failure rate;
- The homogenous process may require a narrow filter to ensure similar equipment, if this approach may be of interest a chi-squared test should be done;
- In case there are several sets of installations and operators, then there is Non-Homogeneous Poisson Process (NHPP) and an Bayesian approach is applicable to obtain a representation for average failure rate and the variability for these data sets;
- Data sets even being NHPP should have consistency amount themselves and should be verified;
- The OREDA estimation should be applied to each failure mode and its Severity Class.

The data extracted from OREDA comes from a certain number of equipment classes, i.e. pumps/centrifugal/oil export, which are composed of identical equipment units. For each class, we have at our disposal: the size of the class, the total number of failures among equipment units of the class and the cumulated operational time for all the equipment units of the class.

For OREDA estimation, the expert should consider a data modelling which will reflect the homogeneity or heterogeneity of failures time series from equipment units.

5. OREDA ESTIMATORS

For now there are two OREDA methods, first method is very basic and does not take into account any heterogeneity between classes, and a second method is elaborated and takes into account heterogeneity. The failure rate estimation is based on both the number of failures and the cumulated operational time among the equipment units of each class.

5.1 OREDA Estimation for HPP

Homogeneous samples are considered those equipments with similar application and power range and within same maintenance procedures. If we face different applications for similar equipment, we should argue and test if those are similar machines.

The first estimation procedure is based on homogeneous samples of an equipment class for same installation and similar maintenance and application. For that, there are pairs (n=failures, t=total time) such: (n_i, t_i), $1 \leq i \leq k$ for k equipments of similar class at the installation.

The solution for this set of homogeneous data consider that each pair (n_i, t_i) has Poisson distribution with a failure rate λ_i with (n_i, t_i) pair. A representative Likelihood function is obtained by the product of a sequence of Poisson distribution for each data set (n_i, t_i) that defines a HPP Homogeneous Poisson Process. For HPP data treatment, on equation 1, failure rates λ remain similar on all subpopulations.

$$L((n_1, t_1), \ldots (n_i, t_i) \mid \lambda) = \prod_{i=1}^{k} \frac{(\lambda\, t_i)^{n_i}}{n_i!} e^{-\lambda\, t_i} \tag{1}$$

The expected value for the failure rate (λ) for all data set of homogeneous sample is the first moment for λ, as equation 2:

$$\hat{\lambda} = E(\lambda) = \int_0^{\infty} \lambda\, L((n_1, t_1), \ldots (n_i, t_i)) d\lambda \tag{2}$$

From the likelihood function (p.d.f.), it is easy to understand that there is a failure rate distribution and from that we need to estimate our trust on the average failure rate. The upper and lower confidence interval (i.e. 5-95% CI) shall be obtained integrating on λ the likelihood function (c.d.f.), as equation 3.

$$\int_{LCI}^{UCI} L(t_1, \ldots t_n) d\lambda = 1 - \epsilon \tag{3}$$

$with$: LCI = Lower Confidence Interval
UCI = Upper Confidence Interval
ϵ = level of significance

On the other hand, Epstein has developed nonparametric estimators for exponential distribution, on equation 4, applicable for determination of confidence intervals.

$$Pr\left[\frac{\chi^2{}_{\epsilon/2}\,(2r)}{2\,T} \leq \hat{\lambda} \leq \frac{\chi^2{}_{(1-\epsilon)/2}\,(2r+2)}{2\,T}\right] = 1 - \epsilon \tag{4}$$

$with$: $r = \sum_{i=1}^{k} n_i$; $T = \sum_{i=1}^{k} t_i$
(n_i, t_i) = equip i[th] of same class and installation
ϵ = level of significance

An example is developed considering OREDA, 5th Ed. Handbook, 2009, taxonomy 1.2.1.3, Gas Turbines Aeroderivative Unknown. Our estimation will reach same values as done on handbook, since there is one installation. It is applied the Poisson Process that end up on exponential distribution and tanking the numbers of failures and cumulative time we can reach on same reliability values as informed on OREDA, see table 1.

Table 1 - Reliability parameters for "Gas Turbines Aeroderivative Unknown"

Population 5	**Installations** 1	**N. of failures**	**Failure rate** (per 10^6 hours)					**Time** (10^6 hours)
Failure Mode			**Lower** (5%)	**Mode** ML(λ)	**Mean** E(λ)	**Upper** (95%)	**SD**	**Calendar**
Critical		55	165,13	209,28	212,73	261,99	26,78	0,2628
								Operational
Critical		55	285,87	362,32	368,28	453,56	46,36	0,1518

Notes: 1. assumed Total Repair Time (Downtime) is not relevant to Total Time
2. average calendar time 52.560 h/equip.
3. average operational time 30.360 h/equip.

5.2 OREDA Estimation for NHPP

The discussion above applies to the situation where the failure rate is treated as a fixed but unknown value, or at least fixed within a small number of subgroups. However, in OREDA database for one type of equipment have been collected failures of a large number of units, so it is not reasonable to assume that each inventory has the same failure rate (λ). In a situation where we have collected data under a variety of operational conditions, it is reasonable to imagine that each inventory has it's unique failure rate.

On OREDA Project most data are from different installations with similar equipment class and for that it is required a Bayesian approach on multi sample condition. The Bayesian claims that the parameter (λ) can never be exactly determined. It would therefore consider (λ) as a random variable, say Λ. A prior probability distribution is postulated reflecting all available information regarding the unknown parameter Λ for which Bayesian inference is desired. The sample data (λ_i) and the prior distribution are combined by Bayes' theorem.

The OREDA estimator does discrete estimation based on Experimental Bayesian (EB) algorithm developed in early 80's. The NHPP is treated with EB model created to represent the average failure rate E(Λ) of all data set inventories selected (population), as represented on equation 7, and the variance V(Λ) between variation of the data sets, as seen on equation 6.

For that, equipment units belonging to the same installation should be merged together but sample sets from different installation have different failure rate (λ_i) that have a failure distribution defined by a Gamma distribution $\Gamma(\alpha, \beta)$. It is estimated first and second moments of failure rate Λ of the multi sample.

An initial estimation of average for Λ is given by:

$$\overline{\Lambda} = \sum_{i=1}^{k} \frac{\sum_{i=1}^{k} n_i}{\sum_{i=1}^{k} t_i} \tag{5}$$

The variance between samples σ_{Λ}^2 is given by:

$$\sigma_{\Lambda}^2 = (V - (k-1)\overline{\Lambda}) \frac{S_1}{S_1^2 - S_1} \tag{6}$$

when greater than 0, else: $\sigma_{\Lambda}^2 = \sum_{i=1}^{k} \frac{\left(\frac{n_i}{t_i} - \overline{\Lambda}\right)^2}{k-1}$

when: $S_1 = \sum_{i=1}^{k} t_i$; $S_2 = \sum_{i=1}^{k} t_i^2$; $V = (V - (k-1)\overline{\Lambda}) \frac{S_1}{S_1^2 - S_1}$

The mean failure rate for Λ is given by:

$$E(\Lambda) = \frac{1}{\sum_{i=1}^{k}\left(\frac{1}{\frac{\overline{\Lambda}}{t_i} + \sigma_{\Lambda}^2}\right)} \sum_{i=1}^{k} \left(\frac{1}{\left(\frac{\overline{\Lambda}}{t_i} + \sigma_{\Lambda}^2\right)} \frac{n_i}{t_i} \right) \tag{7}$$

The failure rate Λ of all set of equipment units for all installations was defined as Gamma Distribution $G(\Lambda; \alpha, \beta)$, alternative form, as defined on equation 8:

$$G(\Lambda; \alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \Lambda^{\alpha - 1} e^{-\Lambda\beta} \tag{8}$$

with : $\Lambda \geq 0$; $\alpha, \beta, \geq 0$

α = shape parameter

β = scale parameter

From that we can associate the mean and variance of failure rate Λ Gamma distribution as that:

$$E(\Lambda) = \frac{\hat{\alpha}}{\hat{\beta}}; \quad V(\Lambda) = \sigma_{\Lambda}^2 = \frac{\hat{\alpha}}{\hat{\beta}^2}$$

The lower and upper confidence interval for failure rate Λ shall be estimated by nonparametric expression for Gamma Distribution, as informed by equation 9, or can be determined by Gamma integration, by equation 10, on its confidence intervals.

$$Pr\left[\frac{\chi^2{}_{\frac{\epsilon}{2}}\,(2\hat{\alpha})}{2\hat{\beta}} \leq \hat{\lambda} \leq \frac{\chi^2{}_{\frac{1-\epsilon}{2}}\,(2\hat{\alpha})}{2\hat{\beta}}\right] = 1-\epsilon \qquad (9)$$

$$\int_{LCI}^{UCI} G(\hat{\alpha},\hat{\beta};\Lambda)d\Lambda = 1-\epsilon \qquad (10)$$

A hypothetical example is given for a set of several samples of similar equipment units within same class but derived from different installations. It is possible to realize the effectiveness of NHPP applied throughout the use of experimental Bayesian algorithm defined on OREDA method. See samples information (Total Time/Failures) and reliability parameters, on table 2, and their distributions, on figure 2.

Table 2 - Samples and NHPP Reliability parameters

Sample	TT (h)	Failures	LCL 5%	Mode ML(λ)	Mean λ	UCL 95%	SD (1/h)	CV (σ/υ)
1	500000	60	9,57E-05	1,20E-04	1,22E-04	1,49E-04	1,47E-05	0,12
2	73000	9	6,43E-05	1,23E-04	1,36E-04	2,15E-04	4,10E-05	0,30
3	73000	2	4,87E-06	2,74E-05	4,02E-05	8,62E-05	2,19E-05	0,55
4	73000	1	7,03E-07	1,37E-05	2,66E-05	6,50E-05	1,77E-05	0,67
HPP	719000	72	8,15E-05	1,00E-04	1,01E-04	1,22E-04	1,12E-05	0,11
NHPP	719000	72	1,60E-05	4,32E-05	7,59E-05	1,72E-04	4,99E-05	0,66

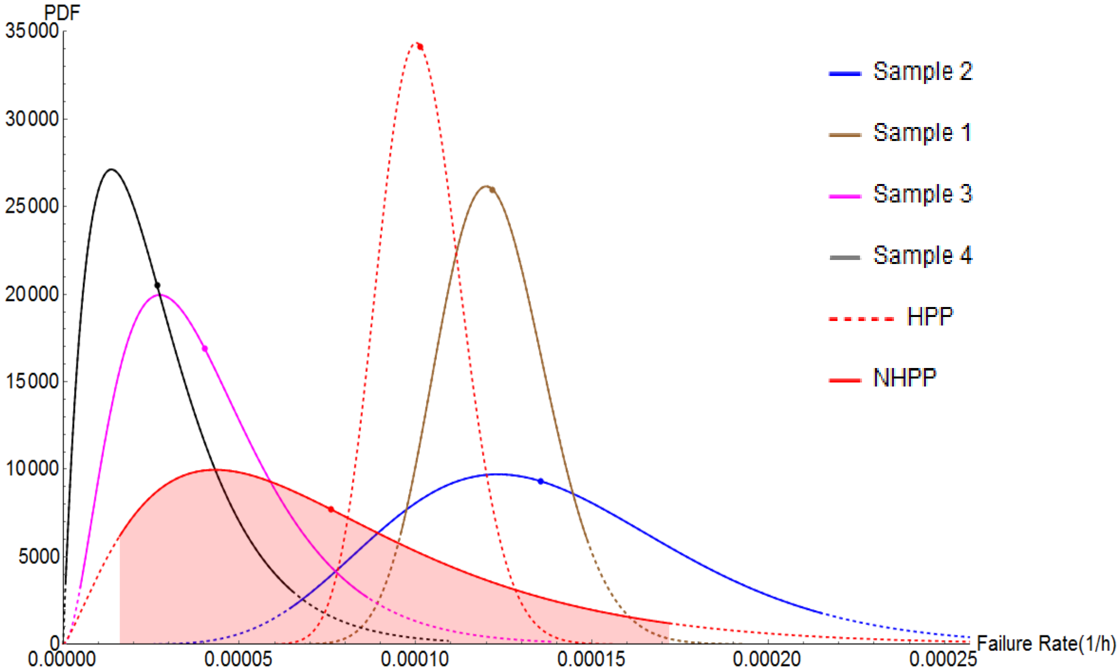

Figure 1 – NHPP vs. HPP distribution

At figure 1, it is observed the benefit to consider correct data treatment for non-homogeneous time series. The mean value on NHPP will represent an average amount each one of all subpopulations averages. On the other hand, a reliability engineer without fundamentals choosing HPP for these time series will get an average value for overall dataset, as homogeneous data.

5.3 Updating a Priori OREDA Information

Throughout Bayesian Law the last informative evidences (likelihood) of an equipment unit can be associated to OREDA priori information and it will be possible to achieve a robust posterior representation for similar units.

The Bayesian approach has to consider the conjugative prior distribution of λ in this case a Gamma Distribution, as equation 11, with known parameters $(\hat{\alpha}, \hat{\beta})$ estimated from OREDA HB, for a specific taxonomy similar to evidence from the plant.

$$G(\lambda \mid \alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\lambda\beta} \tag{11}$$

with : $\lambda \geq 0;\ \alpha, \beta, \geq 0$

The Poisson distribution representing the likelihood, equation 12, of Bayesian approach will consider the n failures observed on k components during the total observation time T.

$$L(n_i, t_i \mid \lambda) = \prod_{i=1}^{k} \frac{(\lambda\, t_i)^{n_i}}{n_i!} e^{-\lambda\, t_i} \tag{12}$$

Appling the Bayesian theory as defined per equation 13.

$$G(\lambda | n_i, t_i) = \frac{L(t_i, n_i|\lambda)\ G_0(\lambda)}{\int_0^{\infty} L(t_i, n_i|\lambda)\ G_0(\lambda)\ d\lambda} \tag{13}$$

The posterior gamma conjugative has analytical solution as presented on equation 14 and indicated on table 3.

$$G(\lambda| \propto +k, \beta + T) = \frac{(\beta + T)^{(\alpha+k)}}{\Gamma(\alpha + k)} \lambda^{(\alpha+k-1)} e^{-\lambda(\beta+T)} \tag{14}$$

Table 3 – Gamma-Poisson Conjugative

Priori $p_0(\theta)$	**Likelihood** $L(x \mid \theta)$	**Posteriori** $p(\theta \mid x)$
$G_0(\lambda \mid \alpha, \beta)$	$P(k \mid T, \lambda)$	$G(\lambda \mid \alpha + k, \beta + T)$

For example: a plant reliability engineer has 3 similar electric motors with ISO 14224 taxonomy A.2.2.4. This set of units, from the same installation, has 2 critical failures and 27.840 h of operation time, and it is required to develop a RCM Plan for these units.

On this case, the reliability engineer does not have too much data from the Asset Maintenance System, so he identified additional data on OREDA Handbook. As a priori information from OREDA, 6th Ed. Handbook, 2015, for Electric Motors, taxonomy 2.2, he found out 91 engines from 10 installations, giving a wide operational time exposure of 1.304.200 h, see table 4.

There are evidences from Bayesian perspective that this likelihood sample (3 electric motors) can be represented by a Poisson distribution and as a priori information there is an OREDA data (91 electric motors) with a very robust information (time exposure) for similar taxonomy.

As presented early, the Experimental Bayesian method applied by OREDA results on Gamma distribution for Λ with data informed on table 4, and is easy to estimate Gamma (α,β) parameters with acceptable error. To estimate the parameters for Gamma curve there is a need to get two points from the Gamma curve, i.e. confidence intervals. Table 4 provides (α,β) parameters and the error estimation for reliability parameters.

Table 4 – Gamma-Poisson Conjugative

<table>
<tr><td>Population</td><td>Installations</td><td></td><td colspan="4">Operational time (10^6 hours)
1,3042</td></tr>
<tr><td>91</td><td>10</td><td rowspan="2">N. of Failures</td><td colspan="4">Failure rate (per 10^6 hours)</td></tr>
<tr><td colspan="2">Failure Mode</td><td>Lower</td><td>Mean</td><td>Upper</td><td>SD</td></tr>
<tr><td colspan="2">Critical</td><td>81</td><td>3,73</td><td>25,04</td><td>62,14</td><td>19,05</td></tr>
<tr><td colspan="2" rowspan="3">Gamma (α,β)
α= 1,75007; β= 69724,6</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td></td><td>3,73</td><td>25,10</td><td>62,14</td><td>19,05</td></tr>
<tr><td>Error</td><td>0%</td><td>0,2%</td><td>0%</td><td>0%</td></tr>
</table>

The Bayesian approach considers independence of distributions as well the non-homogeneous Poisson process. The large uncertainty due to small observed sample of 3 electric motors are anymore a problem since there is a priori information from OREDA handbook. A new posteriori distribution will add non-homogeneous information to the sample and uncertainty is reduced, see on table 5 and its plotting on figure 3.

Table 5 – Gamma-Poisson Conjugative

Sample	LCL 5%	Mode ML(λ)	Mean λ	UCL 95%	SD (1/h)
Priori	3,73E-06	1,08E-05	2,51E-05	6,21E-05	1,90E-05
Likelihood	1,28E-05	7,18E-05	1,05E-04	2,26E-04	5,75E-05
Posterior	1,25E-05	2,82E-05	3,84E-05	7,58E-05	1,98E-05

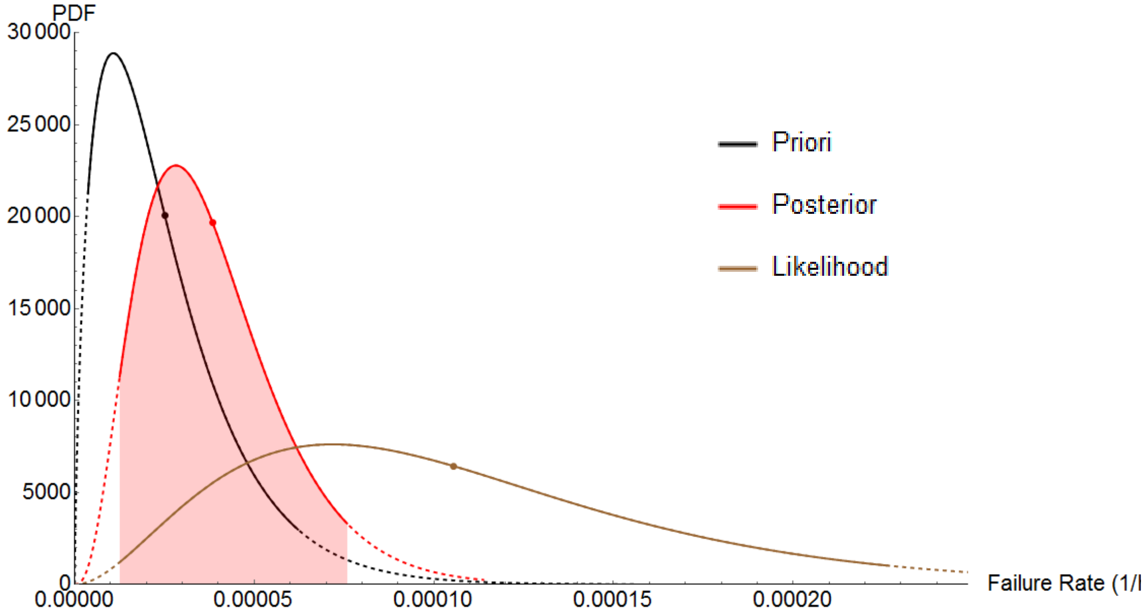

Figure 2 - Updated distribution for FR

The plant engineer now has more information in addition to his small set of electric motors and now he is able to define a better Reliability Center Maintenance RCM Plan for these motors. Since there are just a few failures for his sample, the engineer had thought that the failure rate was a bit high. However, when prior information is added it is foreseen a lower failure rate for these motors.

6. OREDA HANDBOOKS

Each handbook is prepared by SINTEF and NTNU with anonymous data inventory provided by OREDA main contractor from the two last Phases at moment the handbook is built, i.e. Handbook 2015 has data from phases 8 and 9 (project period 2004-2008). Eventually some data collected on early phases can be considered as happen in HB 2015 that collects additional data from phases 6 and 7 (2000-2003). Due to this fact, these handbooks represent early phases from the published data, and unit's faults are not the most recent and may have at least six years old of difference.

The production of handbook comprises data assessment for adequate sampling, delivering estimation for reliability parameters [$E(\Lambda)$,$V(\Lambda)$,$\sigma(\Lambda)$,CI] and maintainability parameters [Mean active Repair Time (MRT), Mean Maintenance Man-hour(MMH)] per severity class and failure modes. SINTEF and NTNU have done these estimations since the 1st handbook 1984.

Handbooks are public edition and everyone may purchase them having access to some equipment taxonomies and major reliability and maintainability parameters. On the other hand, operator's members of OREDA have access to the entire database and can make more smart data filtering and data analysis on the way required for their projects, i.e.: RCM or RAMS analysis.

7. CONCLUSION

OREDA is the oldest JIP in progress, initiated by the Norwegian Petroleum Directorate (now Petroleum Safety Authority, PSA) in 1981. In 1983, several oil companies had joined OREDA and since then 12 phases of data gathering were developed and 6 public handbooks published.

OREDA pioneered discussions on reliability and maintainability database pushing the release of ISO 14224 1st Edition, in July 1999. The formation of the database has long been improved for new equipment taxonomies and so on.

We should follow international standards to keep an acceptable representation of equipment data for reliability and maintainability databank. The quality assurance of the database will give credit to the reliability parameters too.

An overview of OREDA estimators were presented with examples, as well uncertainties were represented since are as important as average values.

Equipments reliability data uncertainties should go along on reliability system estimation resulting in a final value within confidence intervals. For a correct decision process on a reliability study uncertainties should be considered and from that the Asset Integrity Manager will have adequate fundamentals to make a judgment.

8. ACKNOWLEDGMENT

To my college Nicolas Clavé, TOTAL Reliability Engr. and Chairman of OREDA Steering Committee who gently reviewed this paper.

9. REFERENCES:

[1] Mendes, Renato, On the Treatment of Uncertain Data in Reliability Estimation, University of Maryland, 1995;

[2] Martz, Harry F., Bayesian reliability analysis, John Wiley & Sons, 1990;

[3] Mosleh, Ali, Bayesian Methods for Risk and Reliability - Notes for ENRE-640 Draft, University of Maryland, Oct 2007;

[4] OREDA Handbook, 5.Ed., DNV GL, 2009;

[5] OREDA Handbook, 6.Ed., DNV GL, 2015;

[6] Vatn, Jorn, OREDA Data Analysis Guideline, STF75 A93024, SINTEF, 1993;

[7] Vose, David, Risk Analysis-A Quantitative Guide, 3.Ed., Wiley, 2008.