Bulletin of the *Transilvania* University of Braşov
Series I: Engineering Sciences • Vol. 5 (54) No. 2 - 2012

CONDUCTING DATA ANALYSIS FOR ELECTROHYDRAULIC VALVES

C. ŞOIMARU[1] D. BUZEA[1]
S.M. ZAHARIA[1]

Abstract: *The purpose of this work is to determine the life span of the electrohydraulic valve in different test conditions. Using Taguchi method an experiment was designed being formed by combining parameters and their levels. After one test, within the Taguchi test plan, the data obtained needed to be analyzed. Hence, the methodology of analyzing data will be presented so that the medium life span of the valve, in predetermined conditions, could be determined. The first step is to collect experimental data, after its homogeneity must be verified. The third step is to represent graphically experimental data in order to adopt the theoretical distribution law. Using this information, the statistical indicators of the life span will be determined.*

Key words: *valve, experiment, Taguchi, Weibull, life span.*

1. Introduction

In order to make cars friendlier with environment, to obtain less fuel consumption and better performance, internal combustion engine specialists have designed practical solutions to meet requirements above. Thus, one of the solutions is represented by variable valve timing mechanism VarioCam Plus produced by Porsche AG.

Within it the electrohydraulic valve is found, which has a fundamental role obtaining different lifts of the internal combustion valve.

Due to the fact that it functions in the internal combustion engine oil system, after a number of cycles it gets blocked. The root cause that produces valve's blocking it is represented by contaminant present in the oil.

There are several causes for having contaminant in the oil: air filter malfunction, which allows dust to be inducted in the engine; wear of components in relative motion, being cause for having magnetic contaminant in the oil; gums resulted from oil degradation etc.

2. Objectives

What this paper assumes to do is determining life span of the valve in different laboratory conditions which simulate reality.

Thus, after a brainstorming technique, were identified the factors that influence most the life of valve. Among them, four were identified as being the most important: A - contaminant quantity, B - dimension of contaminant, C - Mounting position of the valves, D - type of contaminant.

With these parameters, using Taguchi DOE method, a test plan was designed. It is presented in Table 1.

[1] Dept. D02 - High Tech Products for Automotive Industry, *Transilvania* University of Braşov.

Taguchi Orthogonal Array Table 1

Factors / Nr. of experiment	A	B	C	D
1	**1**	**1**	**1**	**1**
2	1	2	2	2
3	2	1	1	2
4	2	2	2	1
5	3	1	2	1
6	3	2	1	2
7	4	1	2	2
8	4	2	1	1

Factor A involves 4 levels, e.g. four values of contaminant quantity represented by the numbers 1 up to 4. First level 60 grams, second 90 grams, third 120 grams and the forth 150 grams.

Factor B involves only two levels: 45 μm and 150 μm.

Factor C, involves two levels: horizontal and vertical mounting positions.

Factor D, involves two levels: magnetic contaminant and non-magnetic contaminant.

Within the test plan presented in Table 1 only the first experiment was run.

This consists in testing simultaneously four valves from a total number of 28 valves per experiment.

3. Material and Methods

The methodology of data analysis consists in the next steps:

1 - Running experiment on the test bed;
2 - Collecting experimental data;
3 - Verifying statistical homogeneity of experimental data;
3.1 - Verifying the randomness of data;
3.2 - Outliar detection;
4 - Choosing theoretical distribution;
5 - Estimating distribution's parameters;
6 - Validating theoretical distribution, using Goodness-of-fit test;
7 - Accepting/rejecting theoretical distribution;
8 - Estimating reliability parameters of the accepted distribution [2].

Functioning time of the valves Table 2

Test number	Hours	Minutes	Nr. of cycles
1	1:49	109	22890
2	1:45	105	22050
3	1:53	113	23730
4	1:55	115	24150
5	1:33	93	19530
6	1:37	97	20370
7	1:26	86	18060
8	1:30	90	18900
9	1:27	87	18270
10	1:32	92	19320
11	1:40	100	21000
12	1:39	99	20790
13	1:18	78	16380
14	1:15	75	15750
15	1:25	85	17850
16	1:31	91	19110
17	2:15	135	28350
18	1:45	105	22050
19	1:47	107	22470
20	1:40	100	21000
21	1:41	101	21210
22	1:46	106	22260
23	1:40	100	21000
24	1:52	112	23520
25	1:09	69	14490
26	1:29	89	18690
27	1:36	96	20160
28	1:22	82	17220

The first two steps of the methodology are synthesized in Table 2. In order to analyze easier the results, the data was converted from hours into cycles of functioning.

After the experimental data was collected it is necessary to verify statistical homogeneity. First, its randomness must be evaluated. Using bibliographic reference [2] the method named “Criteria length of iteration *K*” was used.

The philosophy of this method consists in:

Defining null hypothesis H_0: the data does not have a random characteristic of distribution.

Defining alternate hypothesis H_1: the data has a random characteristic of distribution.

The numbers of values that compose iteration represent the length of the iteration and it is marked as *K*. Decision regarding statistical hypothesis H_0 is taken considering the next facts:

a) if $K_{max} \le K_{n,\alpha}$ then H_0 is accepted;

b) if $K_{max} > K_{n,\alpha}$ then H_1 is accepted,

K_{max} represents the maximum length of existing iterations; $K_{n,\alpha}$ may be determined with the following formula:

$$K_{n,\alpha} = \frac{\lg \dfrac{-0.43429 \cdot n}{\lg(1-\alpha)}}{\lg 2} - 1, \quad (1)$$

where *n* is number of samples and α represents the confidence level. Consecutive values of analyzed characteristic, which have the same property e.g. they are either bigger or smaller than the median of the data, represents iteration. They are grouped, considering, the median of the data in: bigger values (a), smaller values (b) and equal with median (c) [2].

In Table 3 it may be observed that there exist 2 iterations. M = median is equal to 20580 cycles, n = 28 and α = 0.05, thus $K_{n,\alpha}$ = 8.09 and K_{max} is equal to 14. Roman characters I and II, within Table 3, represent *Iteration length* and *Iteration number*.

Considering condition “b” presented above H_1 is accepted, so data has a random characteristic of distribution. That means the fact only natural causes have influenced the analyzed process, so data is homogeneous [2].

Table 3

Determining randomness of data

Nr. crt.	Nr. of cycles	$x_i > M = a$, $x_i < M = b$, $x_i = M = m$	I	II	Obs.
1	14490	b	1	14	$k_1 = 14$
2	15750	b			
3	16380	b			
4	17220	b			
5	17850	b			
6	18060	b			
7	18270	b			
8	18690	b			
9	18900	b			
10	19110	b			
11	19320	b			
12	19530	b			
13	20160	b			
14	20370	b			
15	20790	a	2	14	$k_2 = 14$
16	21000	a			
17	21000	a			
18	21000	a			
19	21210	a			
20	22050	a			
21	22050	a			
22	22260	a			
23	22470	a			
24	22890	a			
25	23520	a			
26	23730	a			
27	24150	a			
28	28350	a			

The formula used to determine the median M of values within Table 3 is:

$$M = \frac{1}{2} \cdot \left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1} \right). \quad (2)$$

Another step in data analysis and which is enclosed in chapter 2 is outliar detection. In order to detect outliars, Grubbs method was used. The main reason for choosing it is represented by the fact that:

a) it is easy to understand it;

b) allows specification of confidence level [6];

c) it is a test which may be used for all theoretical distributions used in reliability [2].

In order to observe if there are outliars or not, formula (3) will be used:

$$Z = \frac{\left| \bar{x} - x_i \right|}{SD}, \qquad (3)$$

where: Z represents the comparison ratio, $\bar{x}$ - arithmetical mean of the values, x_i current value of number of cycles, *It* represents the abbreviation for iteration and SD standard deviation of data. Using source [6] a critical value Z_{crt} was determined at 2.88, the significance level "α" was chosen to be equal with 0.05 and the final result looks like in Table 4:

Outliar detection Table 4

Nr. crt.	Nr. of cycles	Z
1	22890	0.87
2	22050	0.58
3	23730	1.15
4	24150	1.30
5	19530	0.29
6	20370	0.00
7	18060	0.80
8	18900	0.51
9	18270	0.73
10	19320	0.36
11	21000	0.21
12	20790	0.14
13	16380	1.38
14	15750	1.59
15	17850	0.87
16	19110	0.44
17	28350	2.75
18	22050	0.58
19	22470	0.72
20	21000	0.21
21	21210	0.29
22	22260	0.65
23	21000	0.21
24	23520	1.08
25	14490	2.03
26	18690	0.58
27	20160	0.07
28	17220	1.09

Looking in this table it can be seen that all values of Z are smaller than $Z_{crt,}$ so the conclusion is that there are no outliars in the experimental data.

After it has been proven that there are no outliars, the next step is to chose the theoretical distribution. So, a Weibull distribution is chosen to represent this data. Its probability density function is presented in formula (4):

$$f(T) = \frac{\beta}{\eta} \cdot \left(\frac{T - \gamma}{\eta} \right)^{\beta - 1} \cdot e^{-\left(\frac{T-\gamma}{\eta} \right)^{\beta}}, \qquad (4)$$

where:

$$f(t) \geq 0, T \geq \gamma, \beta > 0, \eta > 0, -\infty < \gamma < \infty .$$

Previous relations are written according [1] and β represents shape parameter, η represents scale parameter, γ represents position parameter and T = cycles of functioning.

The reasons for adopting Weibull distribution are:

a) Due to fact that electrohydraulic valves are part of mechanical domain it is recommended that Weibull distribution should be used [2].

b) Weibull distribution is used in corrosion and wear studies, especially in durability calculus for bearings, tools, gear transmissions, study of mechanical and electrical products endurance, material fatigue - [5].

c) Weibull distribution with $\gamma > 0$ and $\beta > 1$ occur naturally for wear-out situations [1].

d) The life distribution of cycles to failure of solids subjected to fatigue stresses is well represented by Weibull distribution [1].

Further, once it has been adopted the Weibull distribution it is necessary to estimate the parameters, which means determining the estimate value for: β, η and γ. There are two main ways to estimate parameters: a) graphical; b) analytical.

Within Weibull ++7 software was used graphical method named Probability Plotting for a Weibull distribution with two parameters, γ was considered to be zero. Probability Plotting assumes a graphical representation of empirical estimation function $F_n(t_i)$ depending on t_i - functioning time for every sample within the batch with volume 28 components, which is also represented in Figure 1.

If all the points fit to the line it means that there is a Weibull distribution with two parameters and estimated values of β and η are right, so it is not necessary to find the estimated value for the third one, γ, because this is equal to zero. The third parameter of Weibull distribution is used when the data do not fall on a straight line [7]. Looking on Figure 1, it can be observed that the spherical points do not fit the blue line, so γ is needed.

Thus, using the same software and method - Probability Plotting a three parameter Weibull distribution it was used. In the same figure it can be observed that the rectangular black points fit better to the line, which means that exist more exact estimated values of parameters β, η and γ.

In conclusion, Probability Plotting method was used only to know if it is necessary to use a two or three parameter Weibull distribution. Further, analytical method named Maximum Likelihood Estimation (MLE) was used to determine the estimated values of the parameters and it is considered the most robust of the parameter estimation techniques [7].

Basically the MLE method relies on solving the log-likelihood function function of each of the three parameters β, η and γ.

The log-likelihood function has the next mathematical expression:

$$\ln(L) = \Lambda = \sum_{i=1}^{F_e} N_i \cdot \ln\left[\frac{\beta}{\eta}\cdot\left(\frac{T_i-\gamma}{\eta}\right)^{\beta-1}\cdot e^{-\left(\frac{T_i-\gamma}{\eta}\right)^{\beta}}\right] - \sum_{i=1}^{S} N_i' \cdot \left(\frac{T_i'-\gamma}{\eta}\right)^{\beta} + \sum_{i=1}^{F_i} N_i'' \cdot \ln\left[e^{-\left(\frac{T_{Li}''-\gamma}{\eta}\right)^{\beta}} - e^{-\left(\frac{T_R''-\gamma}{\eta}\right)^{\beta}}\right], \quad (5)$$

where: F_e - number of groups of times to failure data group; N_i - the number of times to failure in the i^{th} time to failure data group; β - the shape parameter of Weibull distribution; T_i is the time of the i^{th} group of time to failure data; S is the number of groups of suspension data points; N_i' is the number of suspensions in the i^{th} group of suspension data points; T_i' is the time of the i^{th} suspension data group; F_i is the number of interval failure data groups; N_i" is the number of intervals in the i^{th} group of data intervals; T_{Li}" is the beginning of the i^{th} interval; T_{Ri}" s the ending of the i^{th} interval; γ is the location parameter. Solution of the Equation (5) gives the estimate values of parameters, but not before solving each of the next three equations:

$$\frac{\partial \Lambda}{\partial \beta}=0; \ \frac{\partial \Lambda}{\partial \eta}=0; \ \frac{\partial \Lambda}{\partial \gamma}=0 . \qquad (6)$$

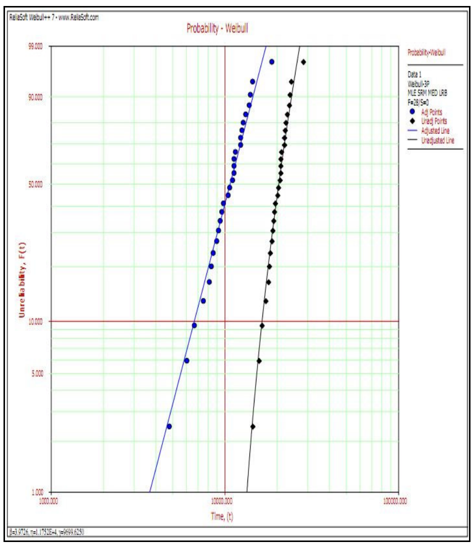

Fig. 1. *Probability Plotting for 3 parameter Weibull distribution*

Using the same software with Equations (4) and (5), the estimated values of β, η and γ with 90% confidence are: β = 3.9726, η = 11752, γ = 9699.625.

After the Weibull distribution was chosen, the natural question that must be put is: does this distribution fits to the experimental data? The answer of this question is given by *Goodness-of-fit tests* that indicate whether or not it is reasonable to assume that a random sample comes from a specific distribution [8].

Now, a test of fit consists in testing a null hypothesis, H_0, which in this case may be defined as:

H_0: the data comes from Weibull CDF (Cumulative Distribution Function) with general form given be Equation (7):

$$F(x,a,b,c)=1-\mathrm{e}^{-\left(\frac{x-a}{b}\right)^{c}} . \qquad (7)$$

It must be specified that experimental data represents complete data.

One of the tests which may answer the question whether the data set can be described by Weibull distribution is the most well-known EDF statistic function Kolmogorov-Smirnov (K-S test) [4]. Kolmogorov-Smirnov function has the main role to measure the minimum distance between the EDF and Weibull CDF defined above.

The EDF function is given by:

$$F_n(x)=\left\{\begin{array}{l} 0, x<X_{1:n} \\ \dfrac{i}{n},\ X_{i:n}\leq x\leq X_{i+1:n}, \qquad i=1\ldots n-1 \\ 1, x\geq X_{n:n} \end{array}\right\} . \qquad (8)$$

The K-S function represents the maximum distance between CDF and EDF [3] and it is calculated as:

$$D=\max\left\{D_n^{+},\ D_n^{-}\right\}, \qquad (9)$$

where: D_n^{+} represents the largest vertical difference when EDF is bigger than CDF; D_n^{-} represents the largest vertical difference when EDF is smaller than CDF [4].

Values of D in excess of the critical value lead to rejection of null hypothesis [3].

With this theoretical background and with the help of Easy Fit 5.5 trial version software, a K-S test was applied to the experimental data obtained on the test bench, Figure 2:

EasyFit - Evaluation Version

Goodness of Fit - Summary

#	Distribution	Kolmogorov Smirnov		Anderson Darling		Chi-Squared	
		Statistic	Rank	Statistic	Rank	Statistic	Rank
1	Weibull (3P)	0.076778	1	0.2561219	1	1.478481	1

Goodness of Fit - Details [hide]

Weibull (3P) [#1]

Kolmogorov-Smirnov

Sample Size Statistic P-Value Rank	28 0.076778 0.9920776 1				
α	0.2	0.1	0.05	0.02	0.01
Critical Value	0.1968	0.22497	0.24993	0.27942	0.29971
Reject?	No	No	No	No	No

Fig. 2. *K-S test using Easy Fit 5.5*

For any confidence levels assigned to the null hypothesis it is clear that the data is correct represented by Weibull distribution, looking on Figure 2.

4. Results and Discussions

For the results box, the last point in the methodology will be presented. Thus, the main reliability characteristics of Weibull distribution are: reliability function, unreliability function, failure rate function. Their formulas are:

- Reliability function:

$$R(T) = e^{-\left(\frac{T-\gamma}{\eta}\right)^{\beta}}; \qquad (10)$$

- Unreliability function:

$$F(T) = 1 - R(T); \qquad (11)$$

- Failure rate function:

$$\lambda(T) = \frac{\beta}{\eta} \cdot \left(\frac{T-\gamma}{\eta}\right)^{\beta-1}. \qquad (12)$$

Using Weibull ++7 the functions have been represented graphically, Figures 3-5.

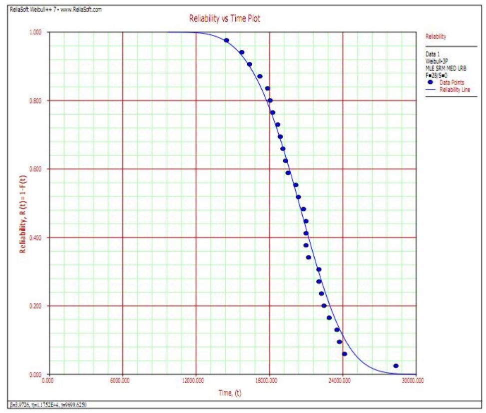

Fig. 3. *Reliability function*

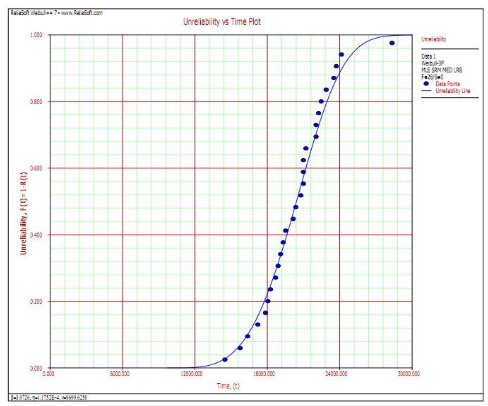

Fig. 4. *Unreliability function*

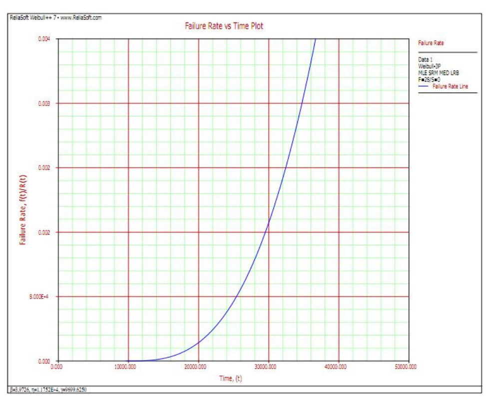

Fig. 5. *Failure rate function*

The Weibull distribution characteristics are also calculated:

1. The Mean of probability distribution function:

$$T_{mean} = \gamma + \eta \cdot \Gamma \cdot \left(\frac{1}{\beta} + 1\right) = 20351.64. \quad (13)$$

2. The Median:

$$T_{median} = \gamma + \eta \cdot \ln(2)^{\frac{1}{\beta}} = 20451.9\,. \quad (14)$$

3. The Mode:

$$T_{\mathrm{mod}e} = \gamma + \eta \cdot \left(1 - \frac{1}{\beta}\right)^{\frac{1}{\beta}} = 20624.34. \quad (15)$$

4. The Standard Deviation:

$$\Gamma_T = \eta \sqrt{\Gamma\left(\frac{2}{\beta} + 1\right) - \left[\Gamma\left(\frac{1}{\beta} + 1\right)\right]^2} = 3228.97. \quad (16)$$

5. Conclusions

Using Design Of Experiment (DOE) Taguchi technique, a test plan was designed according to which 8 experiments are supposed to be run. From the test plan only the first experiment was run and the corresponding data was analyzed so that life span of the valves would be determined.

Using a very clear analysis methodology, the life span of the valves was determined to be equal with 20351.64 cycles which correspond to experiment conditions like: quantity of contaminant was 60 grams with dimensions of approximately 45 µm and magnetic properties having electrohydraulic valves mounted horizontally.

Acknowledgements

This paper is supported by the Sectoral Operational Programme Human Resources Development (SOP HRD), financed from the European Social Fund and by the Romanian Government under the contract number POSDRU/88/1.5/S/59321, POSDRU ID76945.

References

1. Kececioglu, D.: *Reliability Engineering Handbook - Volume I*. Lancaster-Pennsylvania, USA, DEStech Publications, 2002.
2. Martinescu, I., Popescu, I.: *Fiabilitate* (*Reliability*). Braşov, Editura Gryphon, 1995.
3. Murthy, P.: *Weibull Models*. New Jersey, John Wiley & Sons, 2004.
4. Rinne, H.: *The Weibull Distribution - A Handbook*. Giessen, Germany, Taylor and Francis Group, 2009.
5. Zaharia, S.M.: *Cercetări teoretice şi experimentale privind încercarile accelerate de fiabilitate* (*Theoretical and Experimental Research Regarding Realiability Accelerated Tests*). In: Ph.D. Thesis, *Transilvania* University of Braşov, Braşov, Romania, 2010.
6. http://www.graphpad.com. Accessed: 10-01-2012.
7. http://www.reliasoft.com. Accessed: 15-01-2012.
8. http://www.itl.nist.gov. Accessed: 19-01-2012.