Experimental Research Data Series Quality Analysis

Ilmars BLUMBERGS M.Sc., Riga Technology University, Aviation Institute LATVIA; Rafal CHATYS PhD Kielce University of Technology, Faculty of Mechatronics and Machinery Design, POLAND; Prof. Martins KLEINHOFS Riga Technology University, Aviation Institute, LATVIA,

Abstract: *Experiments are necessary part of every scientific work and so much depends on quality of experiments however more or less errors always will be while doing experiments. It is important to have possibility to compare experimental results and to valuate experimental data series, so some kind of result evaluation should be done. To do that simple variant of experimental result evaluation model is offered. In this work some results of single carbon fiber CM thread under static yield force experiments are observed as sample of valuation method. In this work results of two different sample length series are analyzed and presented.*

Keywords: EXPERIMENT RESULT ANALYSIS, QUALITY

1. Introduction

Important part of scientific research work is making and analysis of experiments. In our case we had series of experiments done on our testing equipment (Fig. 1). As a result we had data series with different statistical characteristics. Problem was to determine if our data is acceptable for further calculations, because not only average values were important, but also data distribution character was important.

Fig. 1 Testing equipment with test sample

So to have reliable experimental results to process, some kind of results evaluation should be done even before we start to process our experimental data. Method how to primary verify gained experimental data quality should be fast and easy. It also would be useful to have possibility to compare different quality of experiments of different workers or machines. In this work quality is level of similarity to our expectations.

2. Prerequisites and means for solving the problem

While doing experiments so very many factors have influence on gained results and sometimes it is hard to tell which one factor has greater influence on results and what kind of influence it is. When we dealing with experimental data it means we deal with different kind of statistics and it is important to find what kind of distribution we have and parameters of this distribution, but different experimental environment factors can lead us to wrong decision. By different methods we can find out factors that do influence on test results, but even then we should have some kind of method to decide what data is acceptable and what is unacceptable.

3. Solution of the examined problem

The main idea of offered simple solution is to have such easy method that gives us one number in certain limits that gives us information about processed data quality. This number should be independent from number of statistics or sample dimensions and other values so we easily could compare any data series quality. So simple method, how to qualitatively determine the quality of individual experimental data series in five steps is offered. To determine the quality of a series of experimental data one must:

- determine the data offset from the predicted value;
- determine data dispersion numerically in the percentage part;
- assess the data compliance with the distribution function in accord with Kolomogorov-Smirnov criterion. To get the indicator of interest, divide d2/cn,α;
- visually inspect the distribution graph of the data in the coordinates of interest. If the values are centered and/or distributed in a type of curve, then the first step gets 0 points; if the data is completely chaotic, it gets 0,1, and a neutral rating is 0,005;
- sum the acquired values. That value gained will indicate excellent quality if its close to zero and the data complies with the specified conditions, whereas, if the value exceeds 2, the experimental data is not valid or the corresponding parameters are specified incorrectly.

Those simple steps analyses data from different aspects and covers most characteristics of data series. Each step has its own power and weakest of parameters is visual. All of those steps can be done automatically if software like Excel, Matlab etc. will be used.

4. Results and discussion

To demonstrate this method, results of single carbon fiber composite material thread under static yield force experiments are observed. Those results will be used in strength prediction model, so verification of experimental results, about if they are acceptable for such use, was made. Initial data was stress-strain diagrams and sample of those diagrams we can see in Figure 2.

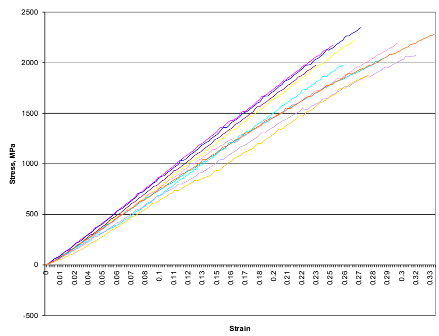

Fig. 2. Sample of initial data for evaluation

From stress-strain diagrams we have a results of ultimate values for two series of 25 tests are shown in Figure 3. Those tests were made on the same equipments for the same material, but with different length. From picture without previous knowledge we can see difference not only in average numbers, but as well difference in dispersion and other characteristics. Goal of experiments was to find different character and tendencies of specimens not absolute numbers by can we say that results are acceptable?

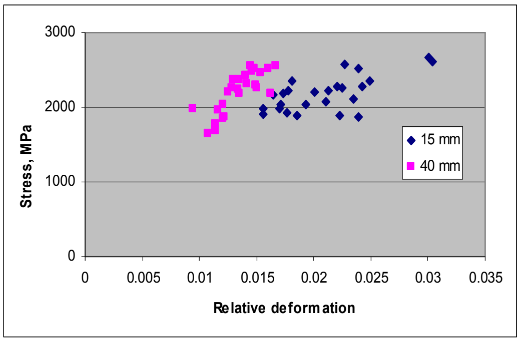

Fig.3 ultimate values for tested samples

So let's start with finding of the data offset from the predicted values. It is easy to find offset by using programs like Matlab, Exel or Statistica. Reviewing the gathered data from a statistical point of view and doing calculations by using expression 1, it can be obtained that the experimental values offset from predicted values for 15 mm sample tensions is k=0,12 and for 40 mm samples – 0,077. Those numbers show us how widely in average data are spread.

$$k = \left(1/n \sum_{i=1}^{n} \sqrt{(a_i - a_{ti})^2} \right) / b$$

n- number of experiments in series;
i – order number;
ai – i-th experimental value;
ati – expected i-th value;
b – expected average value.

Now second easy step that also can be easy done by the same programs is to find data dispersion numerically in the percentage part. It's meant to give us information about extremes. For calculations we have table of basic statistical characteristics for two series of experiments shown in table 1. Looking numerically at the dispersion proportion, composed by dispersion against the average tension value, we have for 15 mm long samples dispersion of 0,1 and for 40 mm – 0,12

Table 1

Basic characteristics of data statistics

	Valid samples	Mean	Minimum	Maximum	Standard deviation
ultimate stress 15mm	25	2181.7	1864.4	2663.7	230.89
ultimate stress 40mm	25	2202.4	1643.9	2554.6	278.81

The third step is to assess the data compliance with the distribution function in accord with Kolomogorov-Smirnov criterion. To get the indicator of interest, divide d2/cn,α; This information will give us number that will characterize similarities of gained data character to expected one. Because this step is more complex and need special tables it is recommended to use program Statistica where it is easy to solve criteria. Model used for determination of criteria is:

$$D_2 = \max_{i=1}^{n} \left\{ \left| \frac{i}{n} - F_x[X_i] \right| \right\}$$

D2 - statistics for n measurements distribution differences from predicted function,
d2 - some value from D2, cn,α defines as P(D2>cn,α)=α,
n - number of experiments,
Xi - i-th measurement

The compliance with normal distribution of the logarithmic values for experimental data was predicted. So tension logarithm distribution compliance with normal distribution for 15 mm and 40 mm long micro samples graphically are shown in Figures 4 a and b, but numerically in the first case we have d2 = 0,088 and cn,α=0,20 and in the second d2 = 0,188 and cn,α=0,20. It means that for our two cases it can be assumed that the experimental data complies with normal distribution; therefore the data can be recognized as valid and for our overall quality criteria it gives values of 0,44 for 15mm samples and 0,94 for 40mm samples.

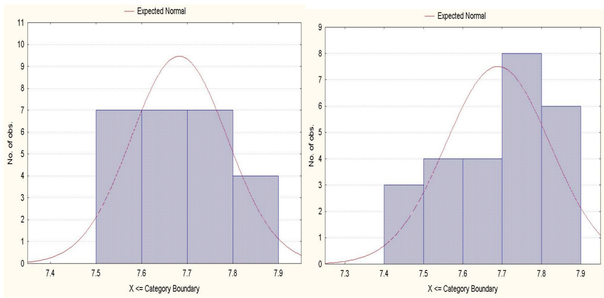

Fig.4 Strength distribution for 15 mm and 40 mm samples

Finally we visually inspect the distribution graph of the data in the coordinates of interest. Because it is subjective measurement this valuation will give us only small influence on overall number. There might be different ways how to depict in best ways goodness of gained data and it depends on what kind of measurement we had. Overall valuation is in invert 10 point scale between 0 and 0,1, where, if the values are centered and/or distributed in a type of curve, then the first step gets 0 points and if data is chaotic it get 10 points and it means 0,1. In our case we can valuate data from figure 3 and figure 5. As we can see data is not chaotic and not perfectly grouped so in both cases we valuated it with 0,05.

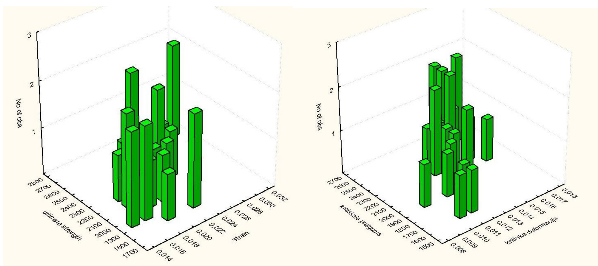

Fig. 5. View of experimental data and its density

As we can see from given sample valuation of data quality is easy to use and understand information of gained value. This method might be improved or specified for some specific data processing. This method approved itself as very useful when we used it to valuate our experimental data that we had obtained from different equipment, different time and people.

5. Conclusions

Simple model for assessing data quality was offered. In example by using two data series was demonstrated how to use offered method. For example, tension values for 15 mm long samples were assessed with 0,71, but 40 mm samples – 1,187, which indicate acceptable quality of gathered data. Such simplified value assessment model allows getting an overview of the experimental data and comparing them. The data series can be assessed as medium good for 15 mm samples and medium for 40mm samples, thus they are valid for use in further research work. Such method can easily be automated can by using this method people or equipment can be monitored for measurement quality and its deviation.

6.References

- J. Paramonovs, J. Andersons, M. Kleinhofs, I.Blumbergs „MINIMAXDM Distribution Family for Analysis of the Tensile Strength of a Unidirectional Composite", Mechanics of Composite Materials, Springer New York, USA, vol 46,No. 3 2010, 275-287 lpp.
- Montgomery, C. Douglas, "Applied Statistics and Probability for Engineers", John Wiley& Sons inc., USA, 2002
- Yu.M Tarnopol'skii, and T.Ya Kincis, "Static Test Methods for Composites", Van Nostrand Reinhold, New York. 1985.
- M. Kleinhofs, „Investigation of Static Strength and Fatigue of Composite Material Used in Aircraft Structure", PhD thesis, Riga. 1983