IMEKO 2010 TC3, TC5 and TC22 Conferences
Metrology in Modern Context
November 22–25, 2010, Pattaya, Chonburi, Thailand

CHARACTERISTIC QUALITY PARAMETERS OF MEASUREMENT - VERIFICATION AND CALIBRATION OF MATERIAL TESTING MACHINES

Dieter Schwenk [1], *Andreas Subaric-Leitis* [2]

[1] Materialprüfungsamt Nordrhein-Westfalen, Dortmund, Germany, schwenk@mpanrw.de
[2] BAM Bundesanstalt für Materialforschung und -prüfung, Berlin, Germany , andreas.subaric-leitis@bam.de

Abstract − The international standards for qualification of material testing machines like the ISO standards for hardness testing [1] - [3] use the relative span (range) as a repeatability parameter. Form the view of statistics and uncertainty the determination of the range is not suitable for this purpose. We propose a transition from range to standard deviation for describing the machines quality parameter which allows a direct estimate of the machines uncertainty and of the reliability of test results within the limits fixed by the ISO standards. For the transition from range to standard deviation the limits have to be transformed. We represent results of conformity decisions determined on the basis of data sets of hardness reference block calibration with reference testing machines using the limits concerning range and the equivalent limits related to the standard deviation.

Keywords: Span, standard deviation, material testing machines

1. INTRODUCTION

The quality parameters of testing or measuring equipment are trueness and repeatability. A measure of the trueness is the deviation of the expectation value of the device (mean) from the true value, a measure of repeatability is the width of the distribution of measurement values, see Fig. 1.

The international standards for calibration and verification of material testing machines use the average of a series of n measurements as an estimate of the expectation and the range as an estimate for the repeatability.

The objective of any measurement is the determination of the expectation value of the measuring device or of a transfer standard and an expectation value for its repeatability. A series of measurements represents a sample used for estimating these parameters. Assuming random straggling of individual measurement values the estimate increases in accuracy as the sampling size increases. In the interpretation of measurements by the Guide to the expression of uncertainty in measurement (GUM) [4] the average represents the best estimate for the true value (if no additional information is available concerning the probable site of the true value). The standard deviation of the measurement series is used for calculating a standard uncertainty or estimating a region in which the expectation value of the measurement device will be found with a particular probability (confidence interval). This provides directly a component of the uncertainty of the device which can be combined with further uncertainty contributions in correspondence with [4]. The range which is applied in international standards is not suited for this purpose. The use of limits for the range has a historical background which no longer applies.

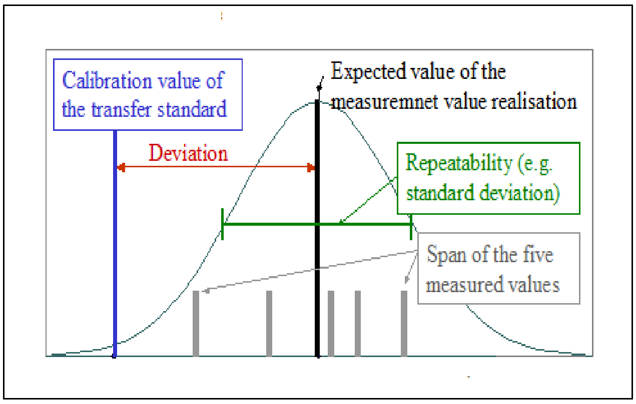

Fig. 1. Evaluation of a measurement series

2. MODDELLING UNCERTAINTY

From the view of metrology the measurand constitutes a random variable. The measurement process represents a sample for estimating the expectation value of the measurand. As mentioned above the value determined by the measurement or test procedure is considered as the best estimate (if no additional information is available) [4]. The true value is assumed to be in the right vicinity of the measurement result with a probability determined by the probability density function. In the case of repeated measurements it is widespread estimating a confidence interval in which the expectation value can be found with the particular confidence level $1\text{-}\alpha$ using the t-distribution. Normal distribution is assumed.

On the other hand, if a standard uncertainty is of interest in the case of repeated measurements, the empirical standard deviation of the sample can be determined by equation (1), which estimates the standard deviation of the random distribution of measurement values. A measure of random straggling is the repeatability, e. g. of a measurement system or a transfer standard. A quantitative value of repeatability can be given e. g. by the standard deviation, by the averaged absolute deviation or by the range determined from the measurement series.

A large number of standards in the field of material testing machines use the range as a measure of straggling of measurement values, i.e. the difference of the maximum and minimum of a measurement series. Assuming a particular distribution of measurement values the probability density function of the range can be calculated and consequently the expectation value of the range. A problem which arises is that the expectation value of the range depends on the number n of values, i.e. on the sample size. Fig. 2 illustrates this situation for samples with sizes $n = 5$, $n = 10$, $n = 50$ assuming normal distribution of measurement values with $\mu = 0$ (mean) $\sigma = 0.25$ (standard deviation). Therefore a comparison of the estimated values of the range determined from different measurements is only possible if the sample size is equal for each sample. On the other hand using the standard deviation its expectation value is robust with respect to the sample size, see Fig. 3.

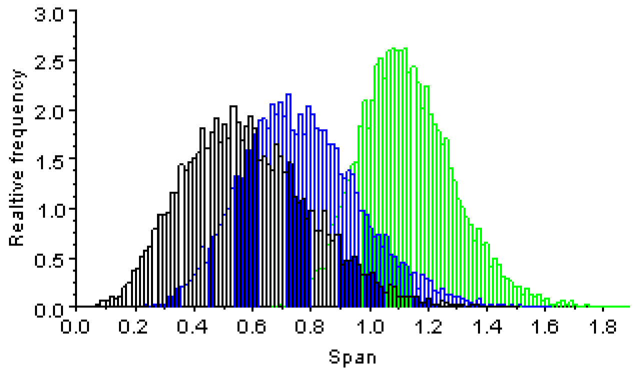

Fig.2 Distribution of the range of a sample ($\mu = 0$; $\sigma = 0.25$, n = 5, n = 10, n = 50), numerical simulation assuming normal distribution of measurement values.

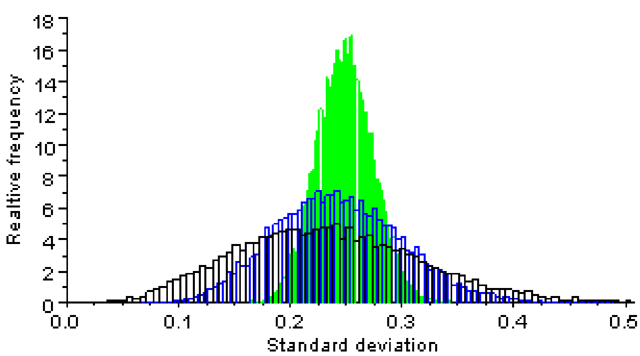

Fig.3 Distribution of the standard deviation of a sample ($\mu = 0$; $\sigma = 0.25$, n = 5, n = 10, n = 50), numerical simulation assuming normal distribution.

Additionally, the range has the following properties: Increasing the sample size the accuracy of the determined value of the range is not increased. A decomposition of straggling data into straggling shares is not possible on the basis of a description using the range. The range does not represent a robust parameter for the description of the straggling of data.

Furthermore, as mentioned above the range is less suitable for the evaluation of uncertainties since standard uncertainties are required which can be estimated by the standard deviations of repeated measurements [4].

Thus we propose the use of standard deviation instead of the range as a measure of repeatability. The advantages are

- The expectation value of the standard deviation concerning the straggling behaviour of the testing equipment and the reference specimen does not depend on the number of repeated measurements.
- The standard deviation (of the population which is related to the straggling) represents a suited measure of repeatability.
- The reliability (confidence level) of the estimate of the standard deviation increases with the increasing number of repeated measurements. The estimate converges to the standard deviation (of the population) „σ".
- Assuming normal distribution of straggling values the probability distribution of the SD of a sample can easily be estimated by the chi-square distribution, the uncertainty of S can be expressed by a confidence interval, see below.
- Analysing the straggling behaviour of data the standard deviation S can be used for variance analysis.

In general the distribution of data follows at least approximately a normal distribution, which is commonly used concerning quality parameters and the evaluation of combined uncertainty. The estimate of standard deviation s is a basic parameter with this respect.

3. DISTRIBUTION OF THE ESTIMATES OF STANDARD DEVIATION

An estimate of the standard deviation of straggling single measurements is (sample size *n*)

$$s(n) = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2} \qquad (1)$$

On condition that the data x_i are normal distributed the confidence interval of the expectation σ of the standard deviation is

$$s^2 \cdot \frac{f}{\chi^2_{f,1-\alpha/2}} \le \sigma^2 \le s^2 \cdot \frac{f}{\chi^2_{f,\alpha/2}} \qquad (2)$$

for a given confidence level $1-\alpha$, (frequently $1-\alpha = 0{,}95$), where $f = n-1$ is the degree of freedom and $\chi^2_{f,1-\alpha/2}$ ($\chi^2_{f,\alpha/2}$) is the argument of the chi-square distribution (quantil) which belongs to the confidence level $1-\alpha/2$ ($\alpha/2$) of the χ^2-

distribution. It is estimated that the expectation σ^2 is within the interval equation (2) with the probability 1-α.

Small sample sizes, that means a small number of single measurements relate to increasing confidence interval equation (2) or in other words the estimate s then represents the expectation σ with increased uncertainty.

4. TRANSFORMATION OF DECISION LIMITS

If the estimate of the standard deviation is used as a measure of repeatability instead of the range the allowed limits, given by the testing machine standards have to be transformed accordingly. This ensures that a testing machine is evaluated and classified nearly equivalently using the standard deviation.

There is the following relation between the standard deviation and the range, where E(R) represents the expectation of the range for a given machine (population)

$$E(R) = \alpha_n \sigma \qquad (3)$$

α_n is the range which is related to the corresponding standard deviation σ and which is generally tabulated in the literature, cf. [5] - [7]. The values of α_n are listed in Tab. 1 for sample sizes n=3 to n=10.

Table 1: Expectation of range ($X_{(n)} - X_{(1)}$) which is related to the standard deviation σ

n	α_n = E(R)/ σ	n	α_n = E(R)/ σ
3	1.7	7	2.70
4	2.06	8	2.85
5	2.33	9	2.97
6	2.53	10	3.08

The limits concerning the estimation of range (denoted as $R_G(n)$) transform to the limits concerning the estimation of standard deviation (denoted here as s_G) accordingly

$$s_G \approx \frac{R_G(n)}{\alpha_n} \qquad (4)$$

or in the case of relative parameters, where V_G denotes the variance coefficient

$$V_G \approx \frac{R_{G\_rel}(n)}{\alpha_n} \qquad (5)$$

Equation (4) and (5) represent estimates of the transformation since $R_G(n)$ and s_G are estimates on the basis of samples.

5. INVESTIGATING THE EQUIVALENCE OF LIMITS USING TEST DATA

Demonstrating the feasibility of the described procedure calibrations of reference hardness blocks have been evaluated as follows

- Transforming the limits given by the hardness measurement standards according to equation (4) or (5), which provides the limits with respect to the variance coefficient.
- Evaluating the data sets and determining the repeatability on the basis of the estimated standard deviation incl. uncertainty.
- Calculation of the relative part of poor hardness blocks (exceeding the limit) and comparing the results when treated on the basis of the range and on the basis of standard deviation (variation coefficient) on the other hand.

Regarding the calibration of hardness blocks according to the relevant standard the percentage of poor reference blocks is compared in view of using the range or using the variation coefficient. Fig. 4 and Fig. 5 show the percentage of rejected reference blocks vs. value of hardness.

The procedures HRC (ISO 6508-3) and HBW10/3000 (ISO 6506-3) have been regarded, the number of repeated measurements or indentations (sample size) is n=5, and the corresponding transformation factor is α_n = E(R)/ σ = 2.33. The limits concerning relative range $R_{G\_rel}(n)$ and V_G are represented in Tab. 2 and Tab. 3.

A number of 6160 at HRC, 3355 at HBW10/3000 and 2766 at HV1 measurement data sets have been evaluated.

Table 2: Relative repeatability $R_{G\_rel}(n)$ on the basis of range and estimates of the transformed limit V_G based on the variation coefficient (n=5) for the remaining indentation depths (HRC).

Remaining indentation depths	Hardness	$R_{G\_rel}(n)$	V_G
> 80 µm		1 %	0,43 %
	> 60 HRC	0,4 HRC	0,17 HRC

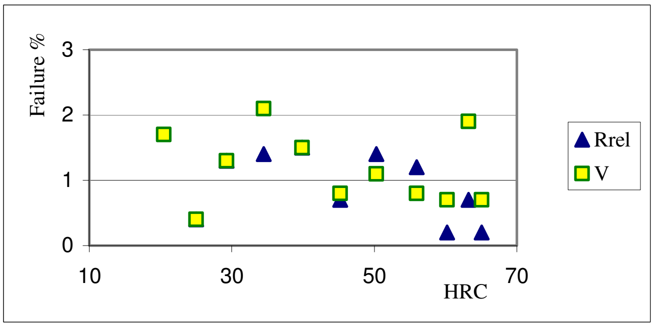

Fig. 4 Calibration of hardness reference blocks (HRC). Percentage of poor reference blocks regarding relative range R_{rel} and variation V coefficient on the other hand. The related limits Tab. 2 have been used for compliance decision.

Corresponding investigations and results are represented in Table 3 and Fig. 5 for the procedure Brinell HBW10/3000 according to ISO 6506-3. n=5, transformation factor α_n = E(R)/ σ = 2.33.

Table 4 and Fig. 6 show the results for the method Vickers HV1 according to ISO 6507-3 with n=5, transformation factor α_n = E(R)/ σ = 2.33

Table 3: Relative repeatability $R_{G\_rel}(n)$ on the basis of range and estimates of the transformed limit V_G based on the variation coefficient (n=5) for the remaining indentation diameter (HBW).

Diameter [mm]	Hardness	$R_{G\_rel}(n)$	V_G
< 0,5	> 200 HB	2 %	0.86 %
0,5 ≤ d ≤1	> 200 HB	1.5 %	0.64 %
> 1	> 200 HB	1 %	0.43 %
	≤ 200 HB	2 %	0.86 %

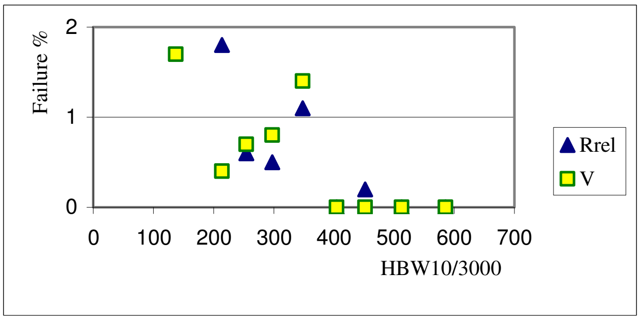

Fig. 5 Calibration of hardness reference blocks (HBW10/3000). Percentage of poor reference blocks regarding relative range R_{rel} and variation V coefficient on the other hand. The related limits Tab. 3 have been used for compliance decision.

Table 4: Relative repeatability $R_{G\_rel}(n)$ on the basis of range and estimates of the transformed limit V_G based on the variation coefficient (n=5) for the remaining diagonals of indentation (HV)

Procedure	Hardness	$R_{G\_rel}(n)$	V_G
HV1	≤ 225 HV1	3 %	1.29 %
HV1	> 225 HV1	2 %	0.86 %

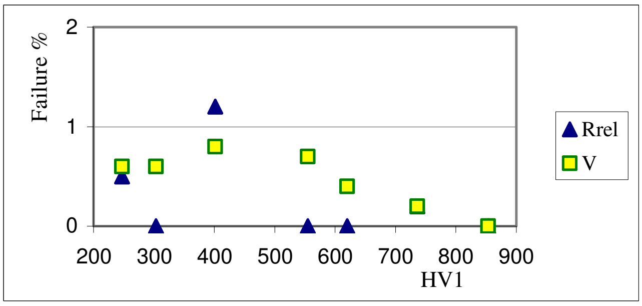

Fig. 6 Calibration of hardness reference blocks (HV1). Percentage of poor reference blocks regarding relative range R_{rel} and variation V coefficient on the other hand. The related limits Tab. 4 have been used for compliance decision.

The represented evaluations concerning the compliance test of hardness reference blocks with respect to the requirements of the repeatability [1] – [3] yield no significant difference of the part of poor items using range or standard deviation (variation coefficient) as a basis of decision. The use of the variation coefficient provides just a marginal higher part of poor items when applying the relation equation (3) for defining the decision limit (transition from range to variation coefficient). In the case of critical straggling of measurements due to outliers the number of measurements can be increased for raising the confidence of the result. This also may reduce the part of items which is tested with negative decision result.

CONCLUSIONS

The range determined by repeated measurements and applied in many standards is less suitable as a measure of repeatability. From the view of statistical treatment and with respect to practical aspects the empirical standard deviation (as an estimate of standard deviation of the straggling of measurement) is the appropriate parameter for evaluation of test items (like material testing machines) and compliance decisions concerning random straggling of measurement data. Using the relation between standard deviation and range the decision limits can be transformed accordingly. The results of investigated data sets of hardness reference block calibrations demonstrate good correspondence of decision results obtained by the variation coefficient and the range which is used in the international standards. The obtained results should also apply to compliance tests / qualification of material testing machines since the basis of decisions and measurements is comparable.

REFERENCES

[1] ISO 6506, Metallic materials- Brinell hardness test- 2 – 3, 2006
[2] ISO 6507, Metallic materials- Vickers hardness test- 2 – 3, 2005
[3] ISO 6508, Metallic materials- Rockwell hardness test- 2 – 3, 2005
[4] Guide to the expression of uncertainty in measurement, ISO, Genf, 1995
[5] H. L. Harter, "Tables of range and studentized range", Ann. Math. Stat., pp. 1122, 1960
[6] L. Sachs, Applied Statistical Methods, Springer, New York, 1999
[7] E. Dietrich, A. Schulze, " Statistical Procedures for Machine and Process Qualification ", Carl Hanser, Muenchen, 2010