Paper

Verification of Staff Proficiency in the Calibration Laboratory on Voltage, Frequency, Resistance and Capacity Measurements

Anna Warzec, Michał Marszalec, and Marzenna Lusawa

National Institute of Telecommunications, Warsaw, Poland

Abstract—The Laboratory of Electrical, Electronic and Optoelectronic Metrology is accredited by the Polish Center of Accreditation for nearly two decades. Over time, the requirements of proficiency verification are growing continuously. In order to satisfy higher and higher of reliability demands the most promising staff proficiency verification estimators were examined for voltage, frequency, resistance and capacity measurements. The article presents measurement systems used for verification, analysis of results or simulations, and shows conclusion for selecting the best solution.

Keywords—Calibration and Measurement Capability, staff proficiency verification.

1. Introduction

The Laboratory of Electrical, Electronic and Optoelectronic Metrology within National Institute of Telecommunications (NIT-LMEEiO) is divided into four divisions (Fig. 1):

- Basic Parameters Metrology Team, which works on the metrology of basic measurements, such as DC&AC, LF voltage and current, resistance, capacitance, inductance, impedance and power,
- Telecommunication Parameters Metrology Team, which works on the measurements of RF and microwave signals and also on transmission parameters of telecommunication networks, e.g. PDH/SDH, Ethernet, SONET etc.,
- Optoelectronic Metrology Team, which works on optoelectronic metrology of such parameters as optical power, wavelength, chromatic and polarization dispersion, optical attenuation and optical fiber length,
- Time and Frequency Metrology Team, which is responsible for accurate measurements of frequency, time, phase time, interval, TIE and conducts science works.

Today's market demands complete offer in electronic area, which NIT-LMEEiO laboratory is trying to fulfill. The work on improvement the quality and CMC (Calibration and Measurement Capability) in every division is continuously ongoing. Every metrology area has its own specifics and in situation of wide range of measured parameters, assurance of proficiency verification and choosing appropriate estimators is not an easy task.

Based on long-term experience the most convenient solution for proficiency verification would be all known and well tested E_n scores. Unfortunately, during the experts discussion and accreditation audits there have been many critical remarks, that for statistically dependent value sets it is not correct solution. In the following discussions, various test and scores have been recommended. In addition, a review of standardization documents [1], [2] (some of them are recently revised) have indicated solutions that could be helpful in solving the proficiency verification problem.

2. Research Plan

For mentioned reasons, following estimators have been analyzed:

- F-Snedecor test [1],
- Bartlett test [1],
- E_n scores [2],
- Morgan test [1],
- ζ scores [2],
- "simple test" – difference of individual measurements should be lower than uncertainty of measurement.

Research was conducted in the following areas of measurements:

- voltage,
- frequency,
- resistance,
- capacity.

In order to detect strong dependence in individual data, for voltage and frequency, the correlation coefficient was calculated for compared pairs of extended result sets (1000 individual measurements). For data set average value, the standard deviation and uncertainty have been calculated. The tables of the critical values for F-Snedecor, Bartlett

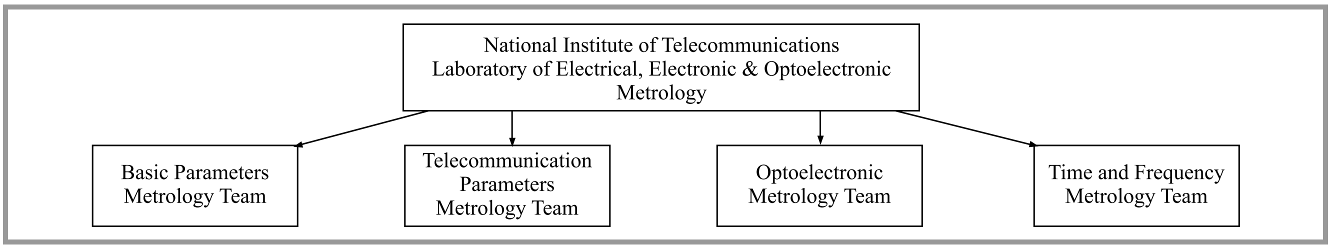

Fig. 1. Actual NIT-LMEEiO laboratory structure.

and Morgan tests have been taken from Excel spreadsheet and verified with statistical tables [1], [3], [4].

2.1. F-Snedecor Test

The F-Snedecor is used to compare standard deviation for two result data sets. It is assumed that both sets have normal distribution. At the beginning standard deviation [5] for both results sets should be calculated. Then F-Snedecor parameter is calculated according to:

$$F = \frac{\dfrac{n_1}{n_1-1}s_1^2}{\dfrac{n_2}{n_2-1}s_2^2}, \tag{1}$$

where for both results sets s_1 and s_2 are standard deviations, n_1 and n_2 are numbers of measurements.

For further analysis, $s_1 > s_2$ constraint must be ensured. The next step is to find appropriate critical value F_{cr} from F-Snedecor distribution table for assumed significance level and calculated degrees of freedom ($f_1 = n_1 - 1$ and $f_2 = n_2 - 1$).

Finally, the comparison between F and F_{cr} has to be made. If calculated value F is lower or equal than F_{cr} ($F \leq F_{cr}$) then conclusion that difference between calculated standard deviation values should not be statistically significant is allowed and the result of proficiency verification is confirmed. In other case ($F > F_{cr}$) the difference is clearly statistically significant and verification of proficiency is unconfirmed.

2.2. Bartlett Test

This test is used to compare standard deviation for many result sets. It is applicable if number of measurements is higher than 3. First, the standard deviation for every result set should be calculated. Then Q parameter is calculated according to:

$$Q = \frac{2.303}{c}\left[(N-k)\log\left(\bar{s}_0^2\right) - \sum_{i=1}^{k}(n_i-1)\log\left(s_i^2\right)\right], \tag{2}$$

where:

$$c = 1 + \frac{1}{3(k-1)}\left(\sum_{i=1}^{k}\frac{1}{n_i-1} - \frac{1}{N-k}\right),$$

$$\bar{s}_0^2 = \frac{1}{N-k}\sum_{i=1}^{k}s_i^2(n_i-1).$$

In presented tests, the same number of measurements in all data sets have been used. Therefore, a simplified following formula is applicable:

$$\bar{s}_0^2 = \frac{1}{k}\sum_{i=1}^{k}s_i^2. \tag{3}$$

In Eqs. (2) and (3): N – summary number of all samples used in calculations, k – number of compared measurement sets, n_i – number of samples in individual data set, s_i – standard deviation for results of method i.

At the end comparison between Q and χ_{kr}^2 has to be made. If calculated value Q is lower or equal than χ_{kr}^2 ($Q \leq \chi_{kr}^2$) then conclusion that the difference between calculated standard deviation values should not be statistically significant is allowed and the result of proficiency verification is confirmed. In other case ($Q > \chi_{kr}^2$) the difference is clearly statistically significant and verification of proficiency is unconfirmed.

2.3. E_n Scores

The E_n scores is mainly used in calibration processes to evaluate inter-laboratory comparisons. The process of evaluation begins of calculation E_n scores according to formula:

$$E_n = \frac{x_i - x_{pt}}{\sqrt{U_i^2 + U_{pt}^2}}, \tag{4}$$

where: x_i and x_{pt} – measured laboratory and reference result values, U_i and U_{pt} – expanded laboratory and reference uncertainties of measured values.

By definition, the final evaluation is done by comparing calculated E_n scores. If $|E_n| < 1$, the evaluation is positive and if $|E_n| \geq 1$ evaluation is negative. Unfortunately this criterion is applicable only if individual data set is statistically independent. In other case scores E_n should be compared to appropriately selected value other than 1. In laboratory practice this critical value is often specified as 0.5 or 0.32.

2.4. Morgan Test

This test is used to compare standard deviation for two correlated result sets.

First, standard deviation for both results sets should be calculated. Then regression parameter r (Pearson product – moment correlation coefficient) is calculated according to:

$$r = \frac{k\sum_{i=1}^{k} x_{1i}\,x_{2i} - \sum_{i=1}^{k} x_{1i}\sum_{i=1}^{k} x_{2i}}{\sqrt{\left[k\sum_{i=1}^{k} x_{1i}^2 - \left(\sum_{i=1}^{k} x_{1i}\right)^2\right]\left[k\sum_{i=1}^{k} x_{2i}^2 - \left(\sum_{i=1}^{k} x_{2i}\right)^2\right]}}\,. \qquad (5)$$

Next, the test parameters L and t should be calculated:

$$L = \frac{4s_1^2 s_2^2(1-r^2)}{(s_1^2+s_2^2)^2 - 4r^2 s_1^2 s_2^2}\,, \qquad (6)$$

$$t = \sqrt{\frac{(1-L)(k-2)}{L}} = \frac{|s_1^2 - s_2^2|}{2s_1 s_2}\sqrt{\frac{k-2}{1-r^2}}\,, \qquad (7)$$

where: k – number of pairs of measurements x_1 and x_2 – individual measurements for compared results sets.

In the next the critical value t_{cr} step is read from Student's t-distribution table for assumed level of significance α (5%) and degree of freedom level to compare calculated t and t_{cr}.

If $t \leq t_{cr}$, then the difference between calculated standard deviation values should not be statistically significant and the verification of staff proficiency is confirmed. In other case ($t > t_{cr}$) the difference between compared values is clearly statistically significant and the staff proficiency is verified negatively.

2.5. *ζ Scores*

ζ (zeta) scores could be useful in proficiency evaluation when the goal is to verify if one participant is able to obtain results close to assigned value within their claimed uncertainty. The second participant's measurement set becomes a source of assigned reference value and standard uncertainty.

$$\zeta = \frac{x_i - x_{pt}}{\sqrt{u_i^2 + u_{pt}^2}}\,, \qquad (8)$$

where: x_i – are participant measured value and $u(x_i)$ – standard uncertainty, x_{pt} and $u(x_{pt})$ have assigned value and standard uncertainty for proficiency testing, e.g. second series of measurements from another participant.

If $|\zeta| \leq 2$ the final evaluation is positive, while $2 < |\zeta| < 3$ the evaluation is doubtful, and if $|\zeta| \geq 3$ evaluation is negative. Unfortunately, those constraints are applicable only if individual data sets are statistically independent. Otherwise, the scores ζ should be compared to appropriately selected value other than 1. In practice this critical limits are specified respectively as 0.64 and 1, i.e. $|\zeta| < 0.64$ result is positive.

It could be noted that for uncertainty coverage factor $k = 2$, which is due to its normal distributions most popular in measurements ζ scores is equivalent to $2E_n$.

2.6. *"Simple test"*

The simple test is based on basic comparison of individual measurements, which difference should be lower than quarter of arithmetically added expanded uncertainties of both sets of measurements.

$$|x_1 - x_2| < \frac{U_1 + U_2}{4}\,. \qquad (9)$$

This procedure could be useful if none of recommended tests proved to be applicable for specific unit of measure proficiency verification.

2.7. *Correlation Coefficient*

In most cases for determination of dependencies in individual measurements series correlation coefficients, the Pearson product – moment correlation coefficient was used. It is also a base for Morgan test calculations and it was determined for the cases where series of data were recorded. The result has been presented in few separate tables to avoid unnecessary misunderstandings.

$$r = \frac{cov(x_1, x_2)}{s_1 s_2}\,, \qquad (10)$$

which is equivalent notation of Eq. (5), where s_1 and s_2 are standard deviations for both results sets.

3. Voltage Measurements

In voltage measurements a precise multimeter was used as the object of study. Three independent calibrators were used as a voltage source (Fig. 2). One of them was used in both cases for AC and DC measurements and the other two were used for AC or DC measurement only. Because the equipment used in research provides an opportunity for automated measurements, at least 1000 samples for set have been saved. For such typical measuring systems in the metrological practice, the behavior of the individual tests for different numbers of samples could be observed. It allows more precise statistical characterization of each data set.

Each series of measurements were conducted independently one after another with both calibrators usage. To

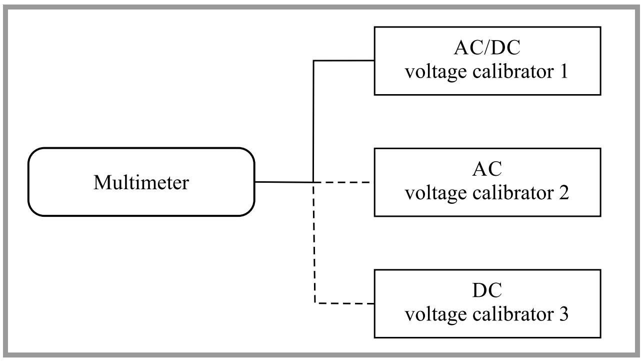

Fig. 2. Measuring system block diagram.

achieve precise statistical analysis the number of individual measurements were set to $N = 1000$. Three different cases were analyzed. The first set was constructed from all of 1000 samples retrieved from both calibrators. In the second case 20 samples from the end of measurement series were used. The last case was set of $N = 10$ samples from the end of measurement series.

3.1. AC Test

The individual series of measurements were saved one after another. The time interval between data sets acquisition was lower than two hours. For measured data, calculations for different numbers of measurements have been done. Figures 3 and 4 show distribution of both result sets while Table 1 presents calculated correlation coefficient for AC voltage measurements.

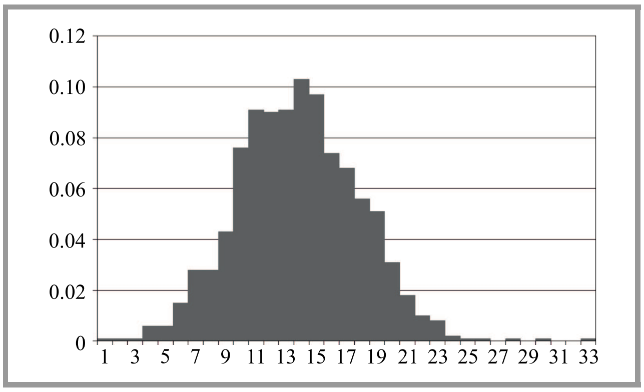

Fig. 3. Distribution graph of the first result set for AC measurements.

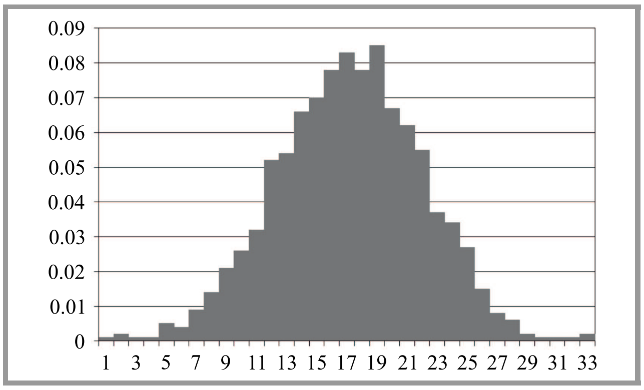

Fig. 4. Distribution graph of the second result set for AC measurements.

Table 1
Correlation coefficients for AC measurement

Number of individual measurements N	Correlation coefficient
1000	3.8%
20	–4.6%
10	12.0%

In Tables 1 and 2 below calculated test results and estimators for N individual measurements have been presented. Value of test parameter: F for F-Snedecor, Q for Bartlett, t for Morgan, $|x_1 - x_2|$ for simple test (ST). Value of estimators: ζ, E_n.

Table 2
Results of analyzed tests for AC voltage measurements

Test name	N	Test result	Critical value	Proficiency verification
F-Snedecor	1000	1.61	1.11	Fail
Bartlett		56.45	3.84	
Morgan		7.62	1.96	
ζ		0.26	0.64	Pass
E_n		0.13	0.32	
ST		0.000046	0.00012	
F-Snedecor	20	2.10	2.17	
Bartlett		2.51	3.84	
Morgan		1.61	2.10	
ζ		0.26	0.64	
E_n		0.13	0.32	
ST		0.000047	0.00012	
F-Snedecor	10	3.00	3.18	
Bartlett		2.50	3.84	
Morgan		1.65	2.31	
ζ		0.25	0.64	
E_n		0.13	0.32	
ST		0.000045	0.00012	

3.2. DC Test

During this measurements sets were also saved one after another. The interval time was lower than two hours. For measured data, calculations for different numbers of measurements have been done. Figures 5 and 6 show distribution of both result sets and Table 3 presents calculated correlation coefficient for DC voltage measurements.

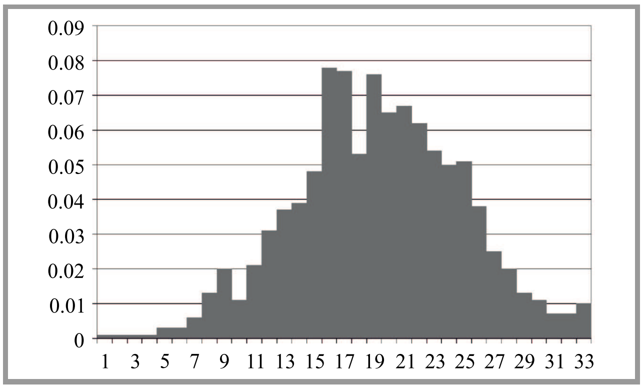

Fig. 5. Distribution graph of the first result set for DC measurements.

In the Table 4 calculated test results and estimators for N individual measurements have been presented.

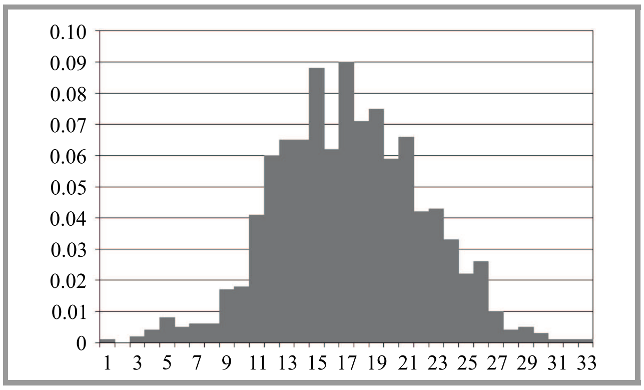

Fig. 6. Distribution graph of the second result set for DC measurements.

Table 3
Correlation coefficients for DC measurement

Number of individual measurements N	Correlation coefficient
1000	6.2%
20	44.1%
10	29.2%

Table 4
Results of conducted tests for DC voltage measurements

Test name	N	Test result	Critical value	Proficiency verification
F-Snedecor	1000	1.13	1.11	Fail
Bartlett		3.92	3.84	
Morgan		1.99	1.97	
ζ		0.17	0.64	Pass
E_n		0.086	0.32	
ST		0.0000022	0.0000090	
F-Snedecor	20	1.18	2.17	
Bartlett		0.13	3.84	
Morgan		0.39	2.10	
ζ		0.21	0.64	
E_n		0.10	0.32	
ST		0.0000026	0.0000090	
F-Snedecor	10	1.10	3.18	
Bartlett		0.021	3.84	
Morgan		0.15	2.31	
ζ		0.22	0.64	
E_n		0.11	0.32	
ST		0.0000028	0.0000090	

4. Frequency Measurements

In the area of frequency measurements a typical signal from internal quartz-driven oscillator was used. Proficiency verification has been analyzed for two persons case. Two independent precise frequency meters synchronized to atomic cesium reference clock were used as a source of standard (Fig. 7). For more precise statistical analysis the number of measurements was set to $N = 1000$ and the gate open time to 1 s. In the research, three different cases were analyzed. To characterize short-term stability of reference oscillator typically the set of 1000 samples is measured with gate open for 1 s. The second case is frequency measurements with assumed normal distribution or with low resolution where many laboratories is using set of 10 samples. For more precise measurements set of $N = 20$ samples is used.

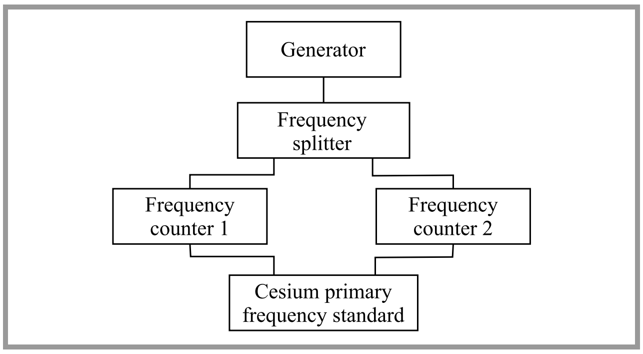

Fig. 7. Block diagram of measuring system.

Because of non-stationary nature of frequency generation process both counters where connected to same source through signal splitter (frequency distribution amplifier)

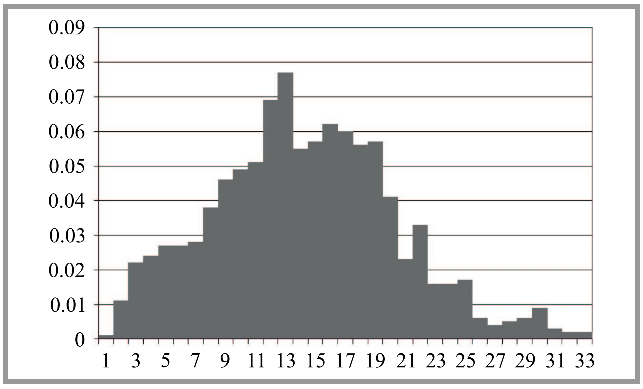

Fig. 8. Distribution graph of the first result set for frequency measurements.

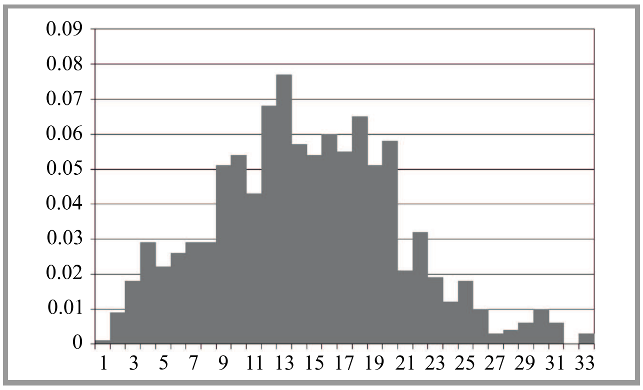

Fig. 9. Distribution graph of the second result set for frequency measurements.

and all measurements were started at the same moment. Figures 8 and 9 show distribution of both result sets and Table 5 presents calculated correlation coefficient for frequency measurements.

Table 5
Correlation coefficients for frequency measurement

Number of measurements N	Correlation coefficient
1000	62.1%
20 (set 1)	99.6%
20 (set 2)	–79.9%
10 (set 1)	96.3%
10 (set 2)	–95.8%

For measured data, calculation for different numbers of measurements have been done. Set 1 was constructed from data from the beginning of measurement series for both participants (for $N = 10$ and $N = 20$). Set 2 was constructed from data from the beginning for one participant and from the end of measurement series for second participant (for $N = 10$ and $N = 20$). The data in this case was expected to be statistically dependent.

In the Table 6 calculated results and estimators for N individual measurements have been presented.

A good example of problems with usage of advanced statistic test is set 1 for $N = 20$ measurements. The data seems to be almost identical, which is confirmed by correlation coefficient calculation (99.6%). It is the biggest obtained correlation coefficient during all tests (Fig. 10). The Morgan test according to theory [1] should be especially useful for statistically dependent data in staff proficiency verification. Unfortunately, it seems it is the only one test that gave a negative result. The research shows that a positive Morgan test result can be obtained, if with an increase of correlation coefficient the difference between standard deviations values is decreasing.

Table 6
Test results for frequency measurements

Test name	N	Test result	Critical value	Proficiency verification
F-Snedecor	1000	1.01	1.11	Pass
Bartlett		0.052	3.84	
Morgan		0.29	1.96	
ζ		0.015	0.64	
E_n		0.0074	0.32	
ST		0.000041	0.0019	
F-Snedecor	20 (set 1)	1.12	2.17	
Bartlett		0.057	3.84	
Morgan		2.61	2.10	Fail
ζ		0.21	0.64	Pass
E_n		0.10	0.32	
ST		0.0013	0.0044	
F-Snedecor	20 (set 2)	2.92	2.17	Fail
Bartlett		5.11	3.84	
Morgan		3.96	2.10	
ζ		5.38	0.64	
E_n		2.69	0.32	
ST		0.046	0.0058	
F-Snedecor	10 (set 1)	1.03	3.18	Pass
Bartlett		0.0019	3.84	
Morgan		0.16	2.31	
ζ		0.12	0.64	
E_n		0.062	0.32	
ST		0.00047	0.0027	
F-Snedecor	10 (set 2)	2.99	3.18	
Bartlett		2.48	3.84	
Morgan		5.68	2.31	Fail
ζ		14.55	0.64	
E_n		7.28	0.32	
ST		0.087	0.0041	

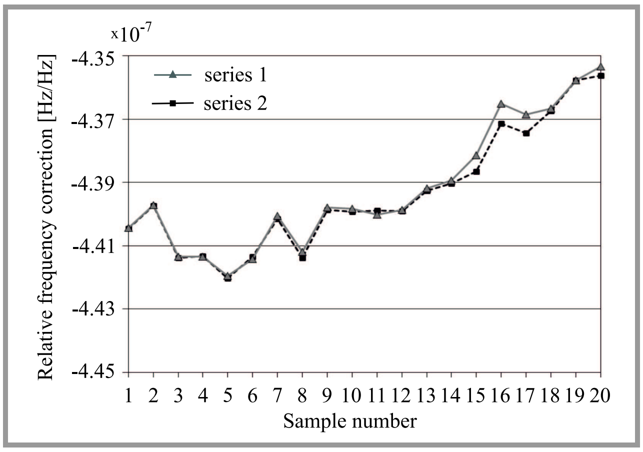

Fig. 10. Results for frequency tests for set 1 and $N = 20$ measurements.

5. Resistance Measurements

In this case the object of study was standard reference resistor. Two series of measurements ($N = 10$ samples) was obtained one after another with usage of two independent precise multimeters (Fig. 11). Authors experience from many years in this area shows that $N = 10$ samples is sufficient to proper characterization of an object, i.e. resistor, because high stability measurements. The limited number of samples and the fact that the dominant component of uncertainty is type B (from the specifications of the instrument), allow to limit the analysis to three last tests (ζ, E_n, ST). Three skipped tests, i.e. F-Snedecor,

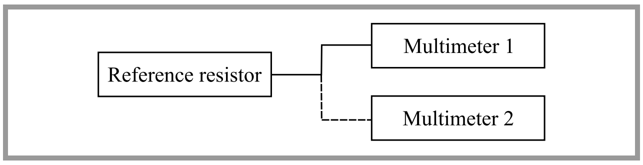

Fig. 11. Block diagram of measurements system.

Bartlett, Morgan, require the usage of a standard deviation (a component of uncertainty of type A).

In the Table 7 calculated test results and estimators have been presented. The set for every value of resistance comes from original measurements.

Table 7
Test results of analyzed tests for resistance measurements

Test name	Value	Test result	Critical value	Proficiency verification
ζ	10 Ω	0.44	0.64	Pass
E_n		0.22	0.32	
ST		0.00006	0.0000925	
ζ	100 Ω	0.045	0.64	
E_n		0.023	0.32	
ST		0.00005	0.000738	
ζ	1 kΩ	0.50	0.64	
E_n		0.25	0.32	
ST		0.00004	0.0000558	
ζ	10 MΩ	0.37	0.64	
E_n		0.19	0.32	
ST		0.00014	0.000245	

6. Capacity Measurements

The object of study (capacity calibrator) was measured on the same measurement station (precise capacity bridge) within the period of four days ($N = 3$ samples) (Fig. 12). The sample number was set to $N = 3$, because measurements process took a long time. For this reason, it seems to be a reasonable choice to characterize the object (capacitor) with good accuracy. As in the previous case, the following discussion was limited to analysis of three tests, i.e. ζ , E_n, ST, due the limited number of measurement samples and the dominant component of uncertainty is type B (information retrieved from the specifications of the instrument).

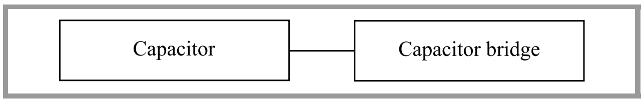

Fig. 12. Block diagram of measurement system.

Table 8 shows calculated test results and estimators. The set for every value of capacity comes from original measurements.

Table 8
Test results for capacity measurements

Test name	Value	Test result	Critical value	Proficiency verification
ζ	10 pF	0	0.64	Pass
E_n		0	0.32	
ST		0	0.00055	
ζ	100 pF	0.13	0.64	
E_n		0.064	0.32	
ST		0.0010	0.0055	
ζ	1000 pF	0.12	0.64	
E_n		0.059	0.32	
ST		0.010	0.060	

7. Conclusions

To determine the most appropriate algorithm for staff proficiency verification few different examples have been tested. First, the voltage measurement were taken one after another, in short (two hours) period of time with usage of two individual high-class reference standards like precise calibrators. The only issue that could affect statistic dependency was the object of study. Due to automation of data acquisition big samples were taken and reliable statistical analysis were done.

The similar process was possible in the frequency case. The measurements were done in parallel at the same time. That is why high correlation factors were expected. Therefore, additional Morgan test was taken under consideration.

The resistance and capacity areas are the examples of typical proficiency verification in the laboratory. Limited number of individual measurements and two different procedures were proceed. For resistance the same object, different precise multimeters and short time interval between measurements was used. For capacity the same object of study and the same measuring equipment was used. However the time between measurements was very large. For those cases the limited number of samples and the fact that the dominant component of uncertainty is type B the analysis was limited to only three tests (ζ, E_n, ST). Three abandoned tests, i.e. F-Snedecor, Bartlett, Morgan require the usage of a standard deviation (a component of uncertainty of type A) which in some cases is not dominant factor of complex uncertainty. For this reason they can be used with particular caution.

As it has been marked, selection one test valid for entire laboratory turns out to be quite difficult. The collected data lead to ambiguous conclusions. In that case, it seems that the most appropriate is to use solutions directly recommended by standardization documents [2], which could be the ζ scores. As it was mention before, for uncertainty coverage factor $k = 2$, which is most popular in measurements due to its normal distributions, ζ scores is equivalent to $2E_n$. Because of statistical dependency, it is suggested to use safer critical limits. In this example limits are specified respectively as 0.64 for ζ.

The good way out of this difficult situation would be usage of proposed simple test or similar equation that matches accuracy level and specificity of individual laboratory.

The results seem to be very promising, but it was worth it to expand the scope of the study in the future. That extension could concern increasing the number of measurements for e.g. capacity (very time-consuming), make study of other physical measures and a comparison between two or more laboratories. That would be the case of interlaboratory comparisons and not only the verification of staff proficiency.

References

[1] P. Konieczka and J. Namieśnik, Eds., *Ocena i Kontrola Jakości Wyników Porównań Analitycznych*. Warsaw, Poland: Wydawnictwa Naukowo Techniczne, 2007 (in Polish).

[2] ISO 13528:2015(E) "Statistical methods for use in proficiency testing by interlaboratory comparison", 2nd ed. ISO 2015, Switzerland 2015.

[3] R. Zieliński, *Tablice Statystyczne*. Warsaw, Poland: Państwowe Wydawnictwa Naukowe, 1972 (in Polish).

[4] A. D. Aczel, *Complete Bussiness Statistics*, 2nd ed. Irwin, Burr Ridge, Illinois; Boston Massachusetts; Sydney, Australia; 1989 and 1993.

[5] JCGM 100:2008, "Evaluation of measurement data – Guide to the expression of uncertainty in measurement", Bureau International des Poids et Mesures, Joint Committee for Guides in Metrology, Sept. 2008 [Online]. Available: www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf

[6] ISO/IEC 17025:2005 General requirements for the competence of testing and calibration laboratories, 2nd ed. Switzerland 2005.

Anna Warzec is a Head of Central Chamber for Telecommunications Measurements and Laboratory of Electrical, Electronic and Optoelectronic Metrology in National Institute of Telecommunications since 1998. Graduation on Faculty of Electrical Engineering of Warsaw University of Technology. She is responsible for supervising the quality management system in the laboratory and on the development of new measurement methods in the area of basic electrical parameters (voltage, resistance, capacity etc.)

E-mail: A.Warzec@itl.waw.pl
Central Chamber for Telecommunications Measurements
National Institute of Telecommunications
Szachowa st 1
04-894 Warsaw, Poland

Michał Marszalec is a Chief of Time and Frequency Metrology Team in Laboratory of Electrical, Electronic and Optoelectronic Metrology in National Institute of Telecommunications. He received the M.Sc. degree in Telecommunications from Faculty of Electronics and Information Technology of Warsaw University of Technology in 2000. He specializes in time and frequency metrology, both measuring techniques and in timescale ensemble algorithms development in Database for Polish Atomic Timescale TA(PL).

E-mail: M.Marszalec@itl.waw.pl
Central Chamber for Telecommunications Measurements
National Institute of Telecommunications
Szachowa st 1
04-894 Warsaw, Poland

Marzenna Lusawa is a Specialist in the Time and Frequency Metrology Team in Laboratory of Electrical, Electronic and Optoelectronic Metrology in National Institute of Telecommunications. She received M.Sc. degree in Faculty of Physics of University of Warsaw in 2008. She specializes in time and frequency metrology, both measuring techniques and in timescale ensemble algorithms development in Database for Polish Atomic Timescale TA(PL).

E-mail: M.Lusawa@itl.waw.pl
Central Chamber for Telecommunications Measurements
National Institute of Telecommunications
Szachowa st 1
04-894 Warsaw, Poland