20th IMEKO TC4 International Symposium and
18th International Workshop on ADC Modelling and Testing
Research on Electric and Electronic Measurement for the Economic Upturn
Benevento, Italy, September 15-17, 2014

# Evaluation of the uncertainty type A of the random stationary signal component from its autocorrelated observations

Z. L. Warsza[1], M. Dorozhovets[2]

[1] *Industrial Research Institute of Automation and Measurement (PIAP)*
*Al. Jerozolimskie 202, 02-486 Warszawa Poland, email: zlw @op.pl*
[2] *Technical University Rzeszow, Poland and Lviv Polytechnic, Ukraine*

***Abstract* – The proposal of evaluating the uncertainty type A of the stationary random component of measured signal from its regularly sampled observations (auto-correlated) is presented. In the first step the regularly variable components of the signal are indentified and removed from the raw sample data. Then upgraded formulas for standard uncertainty type A of the sample and of the mean value are expressed with the use of the correction coefficients or the so-called "effective number" of observations. These quantities depend on number of observations and on the autocorrelation function of the sample cleaned from regular components. Two methods of finding and estimating the autocorrelation function for the sample data are also presented. Few numerical examples are included.**

## I. INTRODUCTION

Most of measured signals have regular and random components. A randomness may exist in a measured object and in a measurement system due to an influence of environment and internal sources. If only some regular components of the signal have to be measured then random component can be considered as the random error and evaluated by the uncertainty type A. In some period of time of collecting the measurement observations this component has the form of random process and can be treated as the stationary one. This random process has the limited frequency range and the observations of a signal are obtained by sampling. If they are very near each other, are autocorrelated. Random deviations from the estimator of measurand, which is the mean value or some regular function, are described by standard deviation of their distribution as the standard uncertainty type A. The international guide GUM [1] gives recommendations how to evaluate in measurement the standard and expanded uncertainties of required probability. The GUM covers only measurement of the variable which data is randomly distributed by Gauss function and their data are not related statistically (i.e. without autocorrelation). Then the use of actual version of GUM [8] is limited in many types of measurements and urgently needs upgrading. In particular for measurements of signals which are variable in time or in space it has been not yet established how to estimate standard deviation (SD) and uncertainties type A. In this paper it is synthetically discussed how to determinate standard deviation and measurement uncertainties of regularly sampled measurand if observations are autocorrelated. In publications [2] - [6] it was enclosed more in detail. It includes the preparation of the raw data sample to further statistical calculations by removal a priori known and unknown regularly variable components [2], [7]. Then the estimation of sample proper standard deviation for known autocorrelation function $\rho$ is given. If this function is unknown, by methods presented in this paper will be able to find the estimator $r$ of the autocorrelation function from the sample data.

## II. THE UNCERTAINTY OF MEAN VALUE FOR CORRELATED OBSERVATIONS

The sequence of measurement data obtained from the sampling process and purified from the deterministic component can be described by a stationary time series. Statistical correlations between realizations $X_i$, $X_{i+k}$ of such series is characterized by the autocorrelation function

$$\rho_k = \frac{\operatorname{cov}(X_i, X_{i+k})}{\sigma^2}. \tag{1}$$

Function $\rho_k$ depends on the frequency spectrum of the test process. Its estimate $r_k$ should be found from the measurement data. In measurements of physical quantities function $\rho_k$ is positive.

The relationship between standard deviation $\sigma(\bar{x})$ of the mean value and $\sigma$ of the individual observation $x_i$ results from variance of the sum of random variables [2 - 4]

$$\sigma(\bar{x}) = \frac{\sigma}{\sqrt{n_{eff}}}, \tag{2}$$

where $$n_{eff} = \frac{n}{1+2\sum_{k=1}^{n-1}(1-k/n)\rho_k} \equiv \frac{n}{1+D_\rho}. \quad (3)$$

For the statistically independent observations $\rho_k \to 0$ (for $k \geq 0$), consequently $D_\rho = 0$ and formula (2) passes to the commonly known relation $\sigma(\bar{x}) = \sigma / \sqrt{n}$ . In opposite, if the observations are fully correlated (closely linked), i.e. $\rho_k \to 1$, with (3a) results

$$D_\rho \to \frac{2}{n}\sum_{k=1}^{n-1}(n-k)\cdot 1 = n-1. \quad (3)$$

Then the standard deviation of the mean is the same as for a single observation because all subsequently repeated observations will be the same in the limit $\rho_k \to 1$. The value of $n_{\mathrm{eff}}$ is needed to proper estimation of the standard deviations $s_a(x_i)$, $s_a(\bar{x})$ for sample of autocorrelated observations [4 - 6]. Their relations to GUM parameters $s(x_i)$, $s(\bar{x}_i)$ are given in Table 1.

Autocorrelation data have the effective number degrees of freedom $\nu_{\mathrm{eff}}$. Defined approximately value of $\nu_{\mathrm{eff}}$ [4-6] is given in Table 1 by eq. (8). In this case $\nu_{\mathrm{eff}} \neq n_{\mathrm{eff}} - 1$ [4, 5]. The relative dispersion of the standard deviation of autocorrelated observation depends on $\nu_{\mathrm{eff}}$. It is also given in Table 1 - eq. (9). Formulas for calculation expanded uncertainties of non-correlated and correlated data are given in the last row of Table 1.

*Table 1. Standard deviation and expanded uncertainty of the sample of non correlated and autocorrelated data.*

| Parameter | Non correlated data (GUM) | Autocorrelated data |
|---|---|---|
| Effective number of observations | $n$ | $n_{eff} = \frac{n}{1+2\sum_{k=1}^{n-1}(1-k/n)r_k} \equiv \frac{n}{1+D_\rho}$ (5)<br>where: $r_k$ - estimation of autocorrelation function in point $k$ |
| Standard deviation of single observation | $s(x_i) = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2}$<br>where: $\bar{x}$ - mean value | $s_a(x_i) = k_a s(x_i)$ (6)<br>where: $k_a = \sqrt{\frac{n_{eff}(n-1)}{n(n_{eff}-1)}} \approx 1$ (6a) |
| Standard deviation of mean value $\bar{x}$ | $s(\bar{x}) = \frac{s(x_i)}{\sqrt{n}} = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n(n-1)}}$<br>$\equiv u_A$ | $s_a(\bar{x}) = \frac{s_a(x_i)}{\sqrt{n_{eff}}} = k_b s(\bar{x}) = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n(n_{eff}-1)}}$ (7)<br>where: $k_b = \sqrt{\frac{n-1}{n_{eff}-1}}$ (7a) |
| Number of freedom | $\nu = n - 1$ | $\nu_{eff} \cong \frac{n}{1+2\sum_{k=1}^{n-1}\rho_k^2} - 1$ (8) |
| Related standard dispersion of standard deviation | $\frac{u(s)}{s} = \frac{u(s(\bar{x}))}{s(\bar{x})} \cong \frac{1}{\sqrt{2\nu}}$ | $\frac{u(s_a)}{s_a} = \frac{u(s_a(\bar{x}))}{s_a(\bar{x})} \cong \frac{1}{\sqrt{2\nu_{eff}}}$ (9) |
| Expanded uncertainty for probability $p$ (coefficient $k_p$ ) | $U = k_p u = k_p\sqrt{u_A^2 + u_B^2}$ | $U = k_p u = k_p\sqrt{s_a^2 + u_B^2}$ (10) |

## III. ESTIMATOR OF THE AUTOCORRELATION FUNCTION OF MEASUREMENT DATA SAMPLE

The autocorrelation function $\rho_k$ is usually not known and needs to be estimated from the measurement data. The most commonly is used form

$$r_k = \frac{\sum_{i=1}^{n-k}(x_i-\bar{x})(x_{i+k}-\bar{x})}{s^2(q_i)}. \quad (11)$$

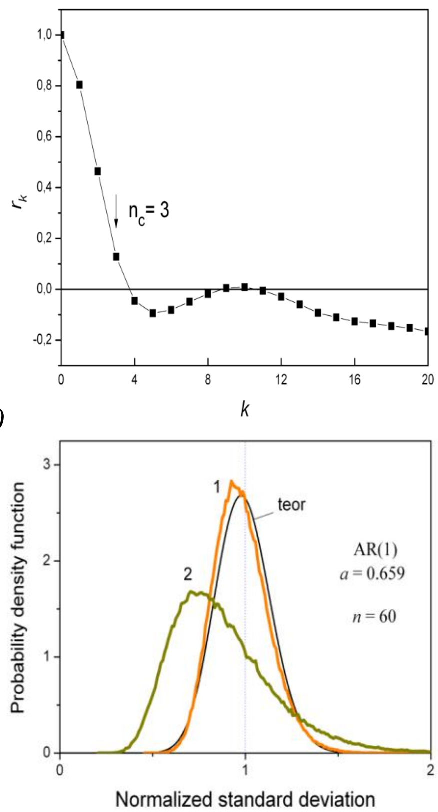
Estimate $r_k$ (Fig. 1a) has two qualitatively different parts. For small $k$ is the falling edge, in which contains real information about the autocorrelation function. The remainder tail is the image of a rather large fluctuations of the correlated noise.

According Zieba [4, 5], the replacement of the function $\rho_k$ in (3) by its estimate $r_k$ is not of the satisfied estimate of the effective number of observations $n_{\mathrm{eff}}$ . The reason is the influence of the tail of autocorrelation function. Zięba proposed to reduce summation in (11) to a few initial estimates of $r_k$ elements, i.e.

$$\hat{n}_{eff} = \frac{n}{1 + 2\sum_{k=1}^{n_c}\left(1 - \frac{k}{n}\right) r_k} \qquad (12)$$

The border value $n_c$ is determined by the last non-zero element of the $r_k$ estimate before its first passage through zero (FTZ method –the first transit through zero). For example, this value for the curve in Fig. 1a is $n_c$=3. FTZ method is valid only for positive correlations.

Figure 1b [6], [4] shows two examples of probability distributions of the estimators obtained by Monte Carlo method. The distribution marked as ‘teor’ in Fig. 1b is calculated from the formula for uncorrelated observations of the standard deviation $z = s/\sigma$ resulting from the distribution of $\chi^2$, where $v$ was substituted by the effective number of degrees of freedom $v_{\text{eff}}$.




*a)*

*b)*

*Fig. 1. a). The initial part of the estimate $\{r_k\}$ of autocorrelation function $\rho_k$ computed from data of Figure 1a (after removing the trend from the raw data); b). Probability density functions for normalized estimators of the standard deviation by the first order autoregressive model AR(1)[1] and the random sample size n=60 [5, 6]. Curves 1 and 2 -derived from MC simulations and relate, respectively, $s_a/\sigma$ and $s_a(\bar{x})/\sigma(\bar{x})$. Curve 'teor' is calculated theoretically by the model (9) for $v_{eff}$ = 22.7.*

[1] Model AR (1) - autoregressive time series of the first order, for which $r_k = a^k$ [4]

Example 1.

Let us find the standard deviation of single measurement $s_a(q_i)$ and of the mean value $s_a(\bar{q})$ of 120 data of observations $q_i$ after withdrawal trend. The estimator $r_k$ of their autocorrelation function is taken for $n_c$=3 according to Figure 1a. The large obtained value $r_k$=0.81 confirms autocorrelation of these data. Standard deviation without considering the autocorrelation is $s(q_i)$=0.0241. At this value of $r_k$ the formula (12) implies the estimate $n_{\text{eff}}$=32.1 and from (6a) and (7a) the coefficient $k_a$=1.012 and $k_b$=1.96. Correlation does not significantly affect the $s_a(q)$=0.0244, and much more the $u_A$, which increases about 2 times from $s(\bar{q}) = 0.00219$ to $s_a(\bar{q}) = 0.00472$. From (9) and (10) is $u(s_a)/s \approx 11\%$.

## IV. APPLICATION OF THE F-TEST FOR INDIRECT DETECTING OF THE AUTOCORRELATION

If number $n$ of observations in the sample of measured process is large enough the original test in [8] is proposed. This test is realized by the separation of $n$ registered observations into $k$ shorter series containing $m$ observations each ($n = m \cdot k$). Then the mean values of these series are estimated and their variances are compared with an average of them for full sample. If the ratio of these variance estimators after normalization by the number of degrees of freedom $k$ - 1 and $n - k$ is greater than the critical value $F_{1-\alpha}(k-1, n-k)$ from the $F$ - distribution (α is a significance level) then it testifies to the auto correlation of observations. The effectiveness of the proposed test is investigated by the Monte-Carlo method.

In order to carry out the indirect identification of autocorrelation the $n$ registered observations are separated into $k$ shorter series of $m$ observations ($m \times k = n$) and registered in a matrix of the size $m \times k$

$$\begin{array}{c} \begin{bmatrix} x_1 & x_{m+1} & \cdots & x_{(j-1)m+1} & \cdots & x_{(k-1)m+1} \\ x_2 & x_{m+2} & \cdots & x_{(j-1)m+2} & \cdots & x_{(k-1)m+2} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_l & x_{l+l} & \cdots & x_{(j-1)m+l} & \cdots & x_{(k-1)m+l} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_m & x_{2m} & \cdots & x_{j\cdot m} & \cdots & x_{k\cdot m} \end{bmatrix} \\ \begin{array}{ccccccc} \bar{x}_{m,1} & \bar{x}_{m,2} & \cdots & \bar{x}_{m,j} & \cdots & \bar{x}_{m,k} & \bar{x}_n \end{array} \end{array} \qquad (13)$$

For each $j$ = 1, 2, .., $k$ (columns) groups of observations in (13) their particular mean values $\bar{x}_{m,j}$ (column) and global mean $\bar{x}_n$ shall be calculated

$$\bar{x}_{m,j} = \frac{1}{m}\sum_{l=1}^{m} x_{(j-1)m+l}\,, \quad \bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{k}\sum_{j=1}^{k}\bar{x}_{m,j} \qquad (14)$$

The estimators of the variations of column $s^2_{m,j}$ and global variation $s^2_n$ are also calculated as:

$$s^2_{m,j} = \frac{1}{m}\sum_{i=(j-1)m+1}^{j\cdot m}\left(x_i - \overline{x}_{m,j}\right)^2, \; s^2_n = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}_n\right)^2, \qquad (15)$$

Using mean of column variations $\overline{s^2_k} = \frac{1}{k}\sum_{j=1}^{k} s^2_{m,j}$ the estimator of the variations of column means $s^2_{\overline{x}_m}$ can be determined as:

$$s^2_{\overline{x}_m} = s^2_n - \overline{s^2_k}\,. \qquad (16)$$

For the normally distributed uncorrelated observations of the variance $\sigma$ according to the estimated by formulas (15) and (16) variances are characterized by the number of degrees of freedom: $s^2_n \Rightarrow n-1$, $\overline{s^2_k} \Rightarrow n-k$, $s^2_{\overline{x}_m} \Rightarrow k-1$. Therefore the random ratio:

$$F_{col} = \frac{s^2_{\overline{x}_m}}{k-1}\Big/\frac{\overline{s^2_k}}{n-k} = \frac{n-k}{k-1}\cdot\left(\frac{s^2_n}{\overline{s^2_k}}-1\right) \qquad (17)$$

is described by the $F$ - distribution with $(n-k,\ k-1)$ degrees of freedom [15]. The expected value and variation of this random ratio accordingly [15] are equals:

$$E[F_{col}] = \frac{n-k}{n-k-2} = 1 + \frac{2}{n-k-2},$$

$$\operatorname{var}[F_{col}] = \left(E[F_{col}]\right)^2\cdot\frac{2\cdot(n-3)}{(k-1)\,(n-k-4)}. \qquad (18)$$

Because $n$ registered observations are located in the natural order in the columns of the matrix (7), the positive autocorrelation of observation causes an increase value of the random ratio of variances $F_{col}$ (17).

For the autocorrelated observations the expectation value of $F_{col}$ depends on two values of indicator $R_\rho(m)$ and $R_\rho(n)$ and also from the number of $n$ observations in $k$ series and it can be calculated by formula:

$$E[F_{col}(\rho)] = \frac{1 + \dfrac{k\cdot R_\rho(m) - R_\rho(n)}{k-1}}{1 - \dfrac{R_\rho(m)}{m-1}}. \qquad (19)$$

Therefore the value of $F_{col}$ can be used for testing the observations autocorrelation. Testing of the absence or presence of the autocorrelation of observation bases on accept or reject the null hypothesis of insignificant or significant variations of mean values across $k$ series (in column) observations. The null hypothesis test conclusion at the specified level of significance α after comparing the observed value $F_{col}$ (14a) with the critical value $F_{critic}(1-\alpha, k-1, n-k)$:

$$F_{col} \le F_{critic}(1-\alpha, k-1, n-k). \qquad (20)$$

Test results:

If $F_{col} \le F_{critic}(1-\alpha, k-1, n-k)$ then **accept** the null hypothesis that variance across $k$ groups observations is insignificant (autocorrelation is insignificant);

If $F_{col} > F_{critic}(1-\alpha, k-1, n-k)$ then **reject** the null hypothesis that variance across $k$ groups observations is insignificant (autocorrelation is significant).

## V. TESTING BY MONTE-CARLO METHOD

The effectiveness of the proposed indirect methods for testing autocorrelation of the observation was carried out by Monte-Carlo method. The normal distributed $N(0,1)$ series of observations of size $n \approx 100$ uncorrelated and correlated data were examined. All observations were divided into $k$ short series of size $m$.

Example 2.

Two autocorrelation functions: exponential (Fig. 1a) $\rho_{ei} = \exp(-a\cdot i)$ (with parameter $a = 0.25$, $0.5$, $0.625,\ldots$ $2.25$) and triangle $\rho_{t\,i} = (1 - i/l)$ for $i \le l$ (with parameters $l = 2, 3, 4,\ldots, 10$) were investigated. The number of simulations $M = 10^4$. The next results mainly relates to the two size $m = 2$ and $m = 5$ of sub series. Therefore the critical values $F_{critic}$ of these cases (level of significance is 0.05) are: $F_q = 1.602$ ($m = 2$) and $F_q = 1.718$ ($m = 5$).

In Fig. 2 two realizations of the uncorrelated ($x_i$, $\rho = 0$) and autocorrelated ($y_i$, $\rho_i = e^{-0.7i}$) observations ($n = 100$) are shown.

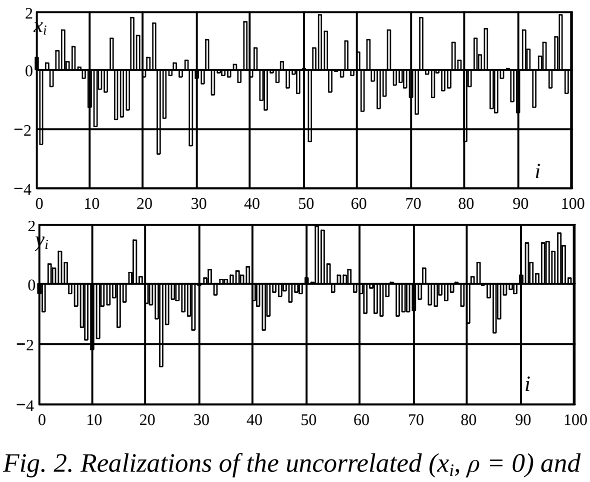



*Fig. 2. Realizations of the uncorrelated ($x_i$, $\rho = 0$) and autocorrelated ($y_i$, $\rho_i = e^{-0.7i}$) observations ($n = 100$).*

Test results of the uncorrelated and correlated observations ($\rho_i = e^{-0.7i}$) are given in Fig. 3a and Fig. 3b. The critical value $F_q$ is the white solid line.

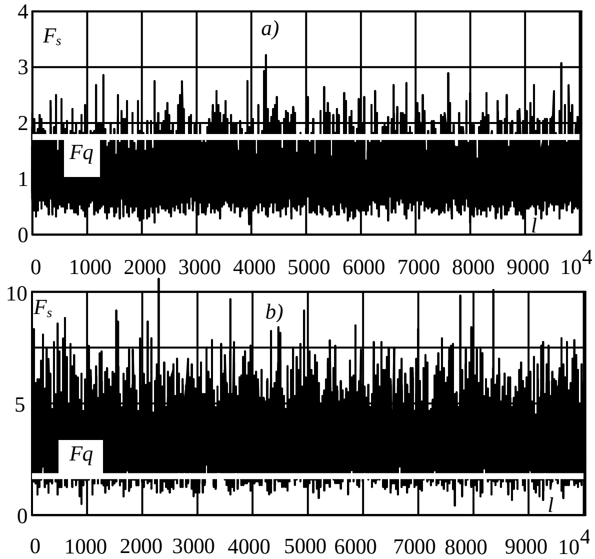



*Fig. 3. M =$10^4$ results of the F - test of n = 100 uncorrelated (a) and autocorrelated ($\rho_i = e^{-0.7i}$) (b) observations determined.*

Analysis results of the statistical simulations research show very good efficiency of the proposed method. Namely, if the autocorrelation of observations is relatively significant where $R_\rho(n) \approx\geq 2$ (for example for the exponential autocorrelation function $\rho_i = e^{-0.7i}$ or triangular $\rho_i = 1 - |i/3|$) then in more than 93% the hypothesis that autocorrelation is insignificant is rejected (the result of test is correct) (Fig. 3).

The probability of the correct correlation testing reaches a maximum value for $m$ = 2 (the shortest series of observations) (Fig. 4). This experimental result is consistent with the theoretical relationship (8) because the maximum value of $E[F_{col}(\rho)]$ is obtained for $m$ = 2.

If the correlation of observations is very small ($R_\rho(n) \to 0$), the hypothesis that autocorrelation is insignificant is accepted (the result of test seeks to correct) and then $n_{eff} = n$. If the correlation theoretically is absent ($R_\rho(n) = 0$) then in 95 % the hypothesis that autocorrelation is insignificant is accepted.

If autocorrelation is more significant ($R_\rho(n) \geq 3$) then more than 99% of the proposed test gives the correct result. Even in the case of relative poorly correlation (only the adjacent 1-2 observations are correlated) where $R_\rho(n) \geq 1$ in approximately 70% (and even more) correlation will be identified correctly (Fig. 4).

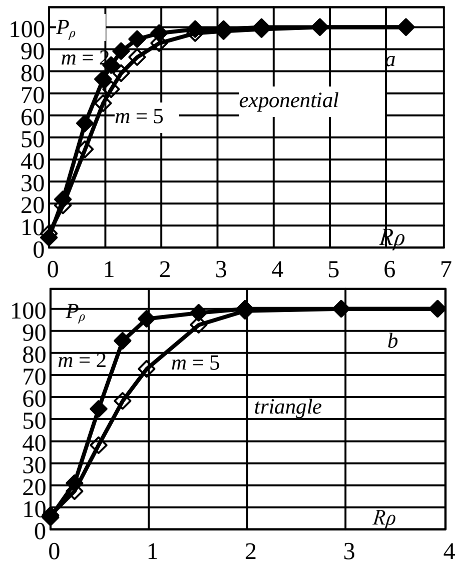



*Fig. 4. Estimated probabilities $P_y$ (%) of the correct identification of correlation versus value of indicator of autocorrelation (m = 2 and m = 5): a – exponential and b – triangle correlation.*

## VI. SUMMARY AND CONCLUSIONS

Discussed issues aroused from the purpose of determining the standard deviation and uncertainty for the measurement data obtained by sampling. This paper summarizes the research results of cleaning such raw data from their regular components [3] and gives the method of calculation the standard deviation for the sample of autocorrelated data. This method is easy to use and allows to extend the range of application of the GUM method type A of estimation uncertainty to such data [1], [2], [4-7].

• Before calculating the standard deviation of the sample of measurement observations as the uncertainty $u_A$ appropriate computational methods must first be used to identify and remove from the raw results the systematic (non periodic and periodic) components.

• For such a randomly distributed values of the observations one needs to know or estimate their autocorrelation function. This function causes a significant increase in uncertainty $u_A(\bar{x})$ compared to the calculated according to GUM. It corresponds to the lower effective number of independent measurements to be taken into account in estimating the standard uncertainty. The content specified adjusted formulas.

• For a limited time to collect measurement observations, a reduction of the sample SD as measurement uncertainty by increasing the sample size (by increasing the sampling rate) is unreliable, as it leads to the necessity to reflect the impact of autocorrelation function of observations.

• Evaluation of the uncertainty $u_A(\bar{x})$ for autocorrelated data discussed here, is valid, as in GUM, for the model of normal distribution. For other distributions this method requires further investigation.

• Programs for the calculation of uncertainty should be supplemented by algorithms for the identification and elimination from the "raw" data the regular components and to obtain estimators of the autocorrelation function from cleaned measured data.

• Presented method allows to upgrade the GUM recommendations by include the proper calculation of the standard and expanded uncertainty for measurand with the stationary random component of their signal.

• The proposed method of testing of observations autocorrelation is realized in a very simple way:

(1) During the registration of $n$ observations the $k$ mean values $\bar{x}_{m,j}$ of $m$ following observations, global mean value $\bar{x}_n$ (14) as well as $k$ column variances $s^2_{m,j}$, their mean value $\overline{s^2_k}$, the global variance $s^2_n$ (15), and variations of column means $s^2_{\bar{x}_m}$ (16) should be calculated;

(2) Normalized to the appropriate numbers degrees of freedom $m$ - 1 and $m – k$ the ratio $F_{col}$ (17) of the variances should be calculated;

(3) This ratio should be compared with the critical value $Fq$ and test result will give an answer about the possibility of the correlation between observations (20).

• From the initial research results can be concluded the high effectiveness of the proposed method of indirect testing of the correlation of observations. If the autocorrelation of observations is relatively significant ($R_\rho(n) \approx\geq 2$) then in more than 93%, the hypothesis that autocorrelation is insignificant is rejected (correct result). If autocorrelation is more significant ($R_\rho(n) \approx\geq 3$) then proposed test gives practically 100 % correct result.

• The result test can be used for the next correction of correlation onto uncertainty of mean value.

REFERENCES

[1] Guide to the Expression of Uncertainty in Measurement, revised and corrected, BIPM 2008

[2] Z. L. Warsza, M. Dorozhovets, M. J. Korczynski, “Methods of upgrading the uncertainty of type A evaluation, Part 1 and Part 2”, Proceedings of 15th IMEKO TC4 Symposium, (2007) Iasi Romania, pp. 193-204

[3] Z. L. Warsza, M. Dorozhovets, “Uncertainty type A evaluation of autocorrelated measurement observations”, Biuletyn WAT (Military Technical Academy) Warszawa 2008, vol. LVII 2, pp. 143-152.

[4] A. Zięba, “Effective number of observations and unbiased estimators of variance for autocorrelated data – an overview”, Metrology & Measurement Systems, 17 (2010), pp. 3-16.

[5] A. Zięba, P. Ramza, “Standard deviation of the mean of autocorrelated observations estimated with the use of the autocorrelation function from the data”, Metrology & Measurement Systems, 18 (2011), pp. 29-25

[6] Z. L. Warsza, “Evaluation of the type A uncertainty in measurements with autocorrelated observations”, Proceedings of jointed TC1+TC7+TC13 IMEKO Symposium 2013, Genua - Journal of Physics: Conference Series 459(2013) 012035 IOP Publishing.

[7] Z. L. Warsza, M. J. Korczynski, “Improving of the type A uncertainty evaluation by refining the measurement data from a priori unknown systematic influences”, Series: Advances in Intelligent Systems and Computing 267 (2014) Springer pp. 727-732.

[8] M. Dorozhovets, “Using F - Test for the Indirect Detecting of the Autocorrelation of Random Observations”, Proceedings of The 7th IEEE Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications. (2013), Berlin, Germany.

[9] J. H. Pollard, “A Handbook of Numerical and Statistical Techniques”, Cambridge University Press, (1977).