

Zygmunt L. WARSZA¹, Mykhaylo DOROZHOVETS^{2,3}

¹INDUSTRIAL RESEARCH INSTITUTE OF AUTOMATION AND MEASUREMENT (PIAP)

202 Jerozolimskie Ave., 02-486 Warszawa, Poland,

²RZESZOW UNIVERSITY OF TECHNOLOGY, 12 Powstańców Warszawy Ave., 35-959 Rzeszów, Poland,

³LVIV POLYTECHNIC NATIONAL UNIVERSITY, 12 Stephan Bandera St., 79013 Lviv, Ukraine

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

Abstract

A proposal of evaluation of the uncertainty type A of the stationary random component of measured signal from its regularly sampled observations when they are auto-correlated is described. In the first step the regularly variable components of the signal are identified 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 use 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 described. Some numerical examples are included.

Keywords: sampled signal, autocorrelated signal, stationary random component, effective number of observations.

1. Introduction

Most of measured signals have regular and random components. Because of influence of environment and internal sources the randomness may exist in the measured object and in the measurement system. 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. This component has the form of random process and can be treated as the stationary one in some period of time of collecting the measurement observations. This random process has the limited frequency range. Then observations of the signal obtained by sampling, if they are very near each other, are autocorrelated. The estimator of measurand is the mean value of some regular function. Random deviations from that function are described by standard deviation of their distribution, and expressed as the standard uncertainty type A. The international guide GUM [1], gives recommendations how to evaluate in measurement the standard and expanded uncertainties with required probability. GUM covers only measurement of variables which randomly distributed data can be described by Gauss probability density function (PDF) and their data are not related statistically (i.e. without autocorrelation). Then 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 was not established yet how to estimate standard deviation (SD) and uncertainties type A.

In this paper it is synthetically discussed how to determine standard deviation and measurement uncertainties of regularly sampled measurand when its observations are autocorrelated. It was presented more in detail in publications [2] - [6]. This includes preparing of the raw data sample to further statistical calculations by removal a priori known and also unknown regularly variable components [2 -Part 1], [7]. Then the proper estimation of standard deviation of the sample for known autocorrelation function ρ is given. In the paper the method is presented, how to find the estimator r of the autocorrelation coefficient from the sample data, if the function ρ is unknown.

2. 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

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

Function ρ_k depends on the frequency spectrum of the test process and is known or its estimate r_k should be found from the measurement data. In measurements of physical quantities function ρ_k is positive.

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

$$\sigma(\bar{x}) = \frac{\sigma}{\sqrt{n_{\text{eff}}}} \quad (2)$$

where:

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

For the statistically independent observations $\rho_k \rightarrow 0$ (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, when the observations are fully correlated (closely linked), i.e. $\rho_k \rightarrow 1$, with (3a) it results

$$D_\rho \rightarrow \frac{2}{n} \sum_{k=1}^{n-1} (n-k) \cdot 1 = n-1 \quad (4)$$

Then the standard deviation of the mean is the same as for a single observation because in the limit $\rho_k \rightarrow 1$ all subsequently repeated observations will be the same.

The value of n_{eff} is needed to properly estimate the standard deviations $s_a(x_i), s_a(\bar{x})$ for sample of autocorrelated observations [4-6]. Their relations to GUM parameters $s(x_i), s(\bar{x})$ are given in the Table 1.

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

3. Estimator of the autocorrelation function of measurement data sample

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

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

Tab. 1. Standard deviation and uncertainty of the sample of non correlated and autocorrelated data

Parameter	Non autocorrelated data (GUM)	Autocorrelated data
Effective number of observations	n	$n_{eff} = \frac{n}{1 + 2 \sum_{k=1}^{n-1} (1-k/n)r_k} \approx \frac{n}{1 + D\rho}$ (5) where: ρ_k - autocorrelation function in point k
Standard deviation of single observation	$s(x_i) = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$ where: \bar{x} - mean value	$s_a(x_i) = k_a s(x_i)$ (6) where: $k_a = \sqrt{\frac{n_{eff}(n-1)}{n(n_{eff}-1)}} \approx 1$ (6a)
Standard deviation of mean value \bar{x}	$s(\bar{x}) = \frac{s(x_i)}{\sqrt{n}} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n(n-1)}}$ $\equiv u_A$	$s_a(\bar{x}) = \frac{s_a(x_i)}{\sqrt{n_{eff}}} = k_b s(\bar{x})$ (7) $= \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n(n_{eff}-1)}}$ where: $k_b = \sqrt{\frac{n-1}{n_{eff}-1}}$ (7a)
Number of freedom	$\nu = n-1$	$\nu_{eff} \approx \frac{n-1}{1 + 2 \sum_{k=1}^{n-1} \rho_k^2} - 1$ (8)
Related standard dispersion of standard deviation	$\frac{u(s)}{s} = \frac{u(s(\bar{x}))}{s(\bar{x})} \approx \frac{1}{\sqrt{2\nu}}$	$\frac{u(s_a)}{s_a} = \frac{u(s_a(\bar{x}))}{s_a(\bar{x})} \approx \frac{1}{\sqrt{2\nu_{eff}}}$ (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)

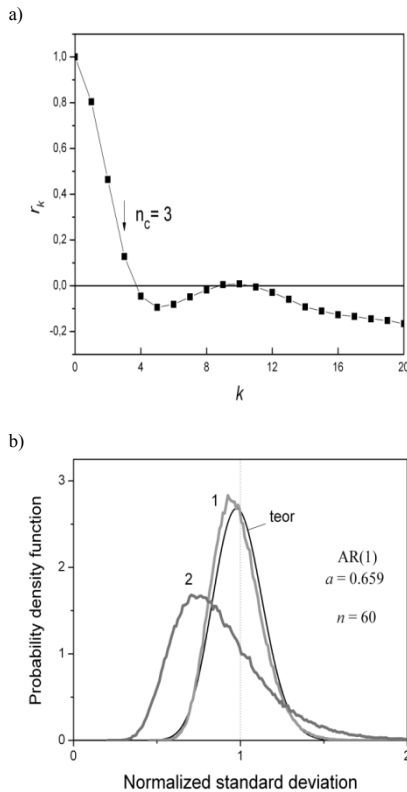


Fig. 1. The initial part of the estimate $\{r_k\}$ of autocorrelation function ρ_k computed from the raw data after removing the trend (a); PDF for normalized estimators of the standard deviation by the first order autoregressive model AR(1) and the random sample size $n=60$ [5, 6] (b). Curves 1 and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

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

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

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

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 given in Figure 1a is $n_c=3$. FTZ method is valid only for positive correlations.

Fig. 1b [6], [4] shows two examples of probability distributions of the estimators obtained by Monte Carlo method. The distribution marked as “teor” in Figure 1b is calculated from the formula for uncorrelated observations of the standard deviation $z=s/\sigma$ resulting from the distribution of χ^2 , where ν was substituted by the effective number of degrees of freedom $\nu_{eff} = n_{eff} - 1$. The model AR (1) there is an autoregressive time series of the first order, for which $r_k = a^k$ [4].

Example

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 simulated observations q_i taken from [7] and given in Figure 2.

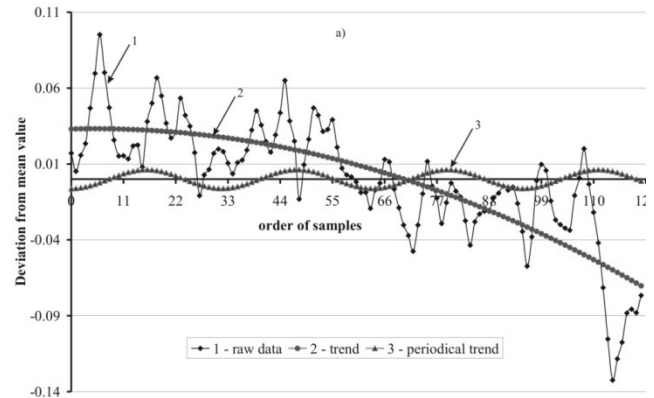


Fig. 2. Sets of measurement data: a) deviations of raw measurement data from mean value in order of the regular sampling and identified systematic components incorporated in data

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 of $r_k=0.81$ confirms the 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 $n_{eff}=32.1$ and from (6a) and (7a) the coefficient $k_a=1.012$ and $k_b=1.96$. Correlation do not affects significantly 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 $u(s_a)/s \approx 11\%$.

4. Using F-test for the indirect detecting of the autocorrelation

Test proposed in [8] is realized by the separation of n registered observations into k shorter series containing m observations ($n = m \cdot k$), and estimation and comparing the variance of the mean values of these series with an average of their variances. 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 autocorrelation of observations.

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

$$\begin{bmatrix} x_1 & x_{m+1} & \dots & x_{(j-1)m+1} & \dots & x_{(k-1)m+1} \\ x_2 & x_{m+2} & \dots & x_{(j-1)m+2} & \dots & x_{(k-1)m+2} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_l & x_{m+l} & \dots & x_{(j-1)m+l} & \dots & x_{(k-1)m+l} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_m & x_{2m} & \dots & x_{j \cdot m} & \dots & x_{k \cdot m} \end{bmatrix} \quad (13)$$

$$[\bar{x}_{m,1} \quad \bar{x}_{m,2} \quad \dots \quad \bar{x}_{m,j} \quad \dots \quad \bar{x}_{m,k}], \quad \bar{x}_n \quad (13a)$$

Mean values $\bar{x}_{m,j}$ for groups of observations of each $j = 1, 2, \dots, k$ column and $l = 1, 2, \dots, m$ row of matrix (13) and the global mean \bar{x}_n are given as (13a). They shall be calculated as

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

The estimators of particular variations of column $s_{m,j}^2$ and the global variation s_n^2 can be calculated as:

$$s_{m,j}^2 = \frac{1}{m} \sum_{i=(j-1)m+1}^{j \cdot m} (x_i - \bar{x}_{m,j})^2, \quad s_n^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x}_n)^2 \quad (15)$$

The estimator of the variations of column means $s_{x_m}^2$ can be determined as:

$$s_{x_m}^2 = s_n^2 - \overline{s_{col}^2} \quad (16)$$

where $\overline{s_{col}^2} = \frac{1}{k} \sum_{j=1}^k s_{m,j}^2$ is a mean of column variations.

For the normally distributed uncorrelated observations of the variance σ^2 according to the estimated by formulas (15) and (16) variances are characterized by the number of degrees of freedom: $s_n^2 \Rightarrow n - 1, \quad s_{x_m}^2 \Rightarrow k - 1, \quad \overline{s_{col}^2} \Rightarrow n - k$. Therefore the random ratio:

$$F_{col} = \frac{s_{x_m}^2 / (k - 1)}{\overline{s_{col}^2} / (n - k)} = \frac{n - k}{k - 1} \cdot \left(\frac{s_n^2}{\overline{s_{col}^2}} - 1 \right) = \frac{n - k}{k - 1} \cdot \frac{1}{s_n^2 / s_{x_m}^2 - 1} \quad (17)$$

are 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:

$$E[F_{col}] = \frac{n - k}{n - k - 2} = 1 + \frac{2}{n - k - 2}, \quad (18)$$

$$\text{var}[F_{col}] = (E[F_{col}])^2 \cdot \frac{2 \cdot (n - 3)}{(k - 1)(n - k - 4)}$$

Because n registered observations are located in the natural order in the columns of the matrix (13), 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 $R_\rho(n)$ (4), and also from the number of m observations in k series and it can be calculated by formula:

$$E[F_{col}(\rho)] \approx \frac{n - k}{n - k - 2} \cdot \frac{1 + \frac{k \cdot D_\rho(m) - D_\rho(n)}{k - 1}}{1 - \frac{D_\rho(m)}{m - 1}}. \quad (19)$$

Therefore the value of F_{col} can be used for testing the observation 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} (17) with the critical value $F_{critic}(1 - \alpha, k - 1, n - k)$:

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

Test results:

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

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

The effectiveness of the proposed indirect method for testing autocorrelation of the observation carried out by Monte-Carlo method and conclusions are given in [10]. Obtained from Monte-Carlo simulation the probability of the correct identification of correlation versus value of indicator $D_\rho(m)$ is presented in Figure 3.

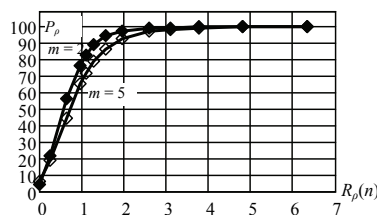


Fig. 3. Estimated probabilities P_p (%) of the correct identification of correlation versus value of indicator $D_\rho(n)$ of autocorrelation $\rho = e^{-0.7t}$ ($m = 2$ and $m = 5$) for $M = 10^4$ Monte-Carlo simulations and $n = 100$ observations

The maximal probability of the correct identification of correlation can be obtained for $m = 2$ (the shortest series of observations, that is consistent with the theoretical relationship (18) because the maximum value of $E[F_{col}]$ in (18) is obtained for $m = 2$.

If the autocorrelation of observations is relatively significant, where $D_p(n) \geq 2$ for example for the autocorrelation, then in more than 93% and if $D_p(n) \geq 3$ then in more than 99% the hypothesis that autocorrelation is insignificant is rejected (the result of test is correct) the hypothesis that autocorrelation is insignificant is rejected (the result of test is correct) (Fig. 3).

Presented methods 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.

5. References

- [1] Guide to the Expression of Uncertainty in Measurement, revised and corrected. BIPM JCGM 100:2008.
- [2] Warsza Z. L., Dorozhovets M., Korczynski M. J.: Methods of upgrading the uncertainty of type A evaluation, Part 1 and Part 2, Proceedings of 15th IMEKO TC4 Symposium, 2007, Iasi Romania, 193-204.
- [3] Warsza Z. L., Dorozhovets M.: Uncertainty type A evaluation of autocorrelated measurement observations. Biuletyn WAT (Military Technical Academy) Warszawa, 2008, vol. LVII 2 143-152.
- [4] Zięba A.: Effective number of observations and unbiased estimators of variance for autocorrelated data – an overview. Metrology & Measurement Systems, 2010, no 17, pp. 3-16.
- [5] Zięba A., Ramza P.: Standard deviation of the mean of autocorrelated observations estimated with the use of the autocorrelation function from the data. Metrology & Measurement Systems, 2011, no 18, pp. 29-25.
- [6] Warsza Z. L.: Evaluation of the type A uncertainty in measurements with autocorrelated observations. CD Proceedings of jointed TC1+TC7+TC13 IMEKO Symposium. 2013, Genua - Journal of Physics: Conference Series 459, 012035 IOP Publishing.
- [7] Warsza Z. L., Korczynski M. J.: 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, no 267 (2014), Springer, pp. 727-732.
- [8] Dorozhovets M.: 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.

- [9] Pollard J. H.: A Handbook of Numerical and Statistical Techniques. Cambridge University Press, 1977.
- [10] Warsza Z. L., Dorozhovets M.: Evaluation of the uncertainty type A of the random stationary signal component from its autocorrelated observations. Proceedings of 20th IMEKO TC4 International Symposium. Benevento, Italy, Sept.15-17, (2014), ISBN-14: 978-92-990073-2-7 p. 367 - 372 + ACTA IMEKO

Received: 17.05.2015

Paper reviewed

Accepted: 02.07.2015

Prof. Mykhaylo DOROZHOVETS, DSc, eng.

Graduate (1975) at the Department of Information and Measurement Technology, candidate of technical sciences (PhD) in 1986 and defended Doctor of technical science (habilitation) in 2001. He is now professor in the Department of Metrology and Measurement Systems of Rzeszow University of Technology and the Lviv Polytechnic National University, Ukraine. He conducts research in the field of tomography, the measurement and signal processing, analysis and evaluation of the uncertainty of measurement results.

e-mail: michdor@prz.edu.pl



Assoc. Prof. Zygmunt L. WARSZA, PhD, eng.

Born in Wilno. Graduated at Electrical Faculty of Warsaw Technical University: MSc. El. Eng. 1959. PhD in El. Measurements 1967. Assoc. Prof. (docent) from 1970. Worked in 3 Technical Universities and 3 research institutes 1959 -2002. Presently with Industrial Research Institute of Automation and Measurements (PIAP) Warszawa. President of Polish Society of Metrology. Author or co-author above 220 papers, 4 monographs, many reports and 11 patents on new instruments.

e-mail: zlw@op.pl

