Metrol. Meas. Syst., Vol. XIX (2012), No. 4, pp. 787-796.

PAN POLSKA AKADEMIA NAUK

METROLOGY AND MEASUREMENT SYSTEMS

Index 330930, ISSN 0860-8229

www.metrology.pg.gda.pl

M&MS

STANDARD UNCERTAINTY DETERMINATION OF THE MEAN FOR CORRELATED DATA USING CONDITIONAL AVERAGING

Adam Kowalczyk, Anna Szlachta, Robert Hanus

Rzeszow University of Technology, Department of Metrology and Diagnostic Systems, Powstańców Warszawy 12, 35-959 Rzeszow, Poland (kowadam@prz.edu.pl, ✉ annasz@prz.edu.pl , +48 17 865 1575, rohan@prz.edu.pl)

Abstract

The correlation of data contained in a series of signal sample values makes the estimation of the statistical characteristics describing such a random sample difficult. The positive correlation of data increases the arithmetic mean variance in relation to the series of uncorrelated results. If the normalized autocorrelation function of the positively correlated observations and their variance are known, then the effect of the correlation can be taken into consideration in the estimation process computationally. A significant hindrance to the assessment of the estimation process appears when the autocorrelation function is unknown. This study describes an application of the conditional averaging of the positively correlated data with the Gaussian distribution for the assessment of the correlation of an observation series, and the determination of the standard uncertainty of the arithmetic mean. The method presented here can be particularly useful for high values of correlation (when the value of the normalized autocorrelation function is higher than 0.5), and for the number of data higher than 50. In the paper the results of theoretical research are presented, as well as those of the selected experiments of the processing and analysis of physical signals.

Keywords: uncertainty of the mean value, autocorrelated data, conditional averaging, random signal.

1. Introduction

In the digital methods of calculations of random signal characteristics or those determined and interfered with random impact occurrences, the accuracy of the estimates of the parameters being calculated depends on the number of data (of quantizated signal samples). Within the given analysis time an increase of the data number makes lower the component of the estimator variance which is data number dependent, and at the same time it increases the variance component dependent upon the higher correlation of the samples from the location situated in smaller distance (in time, impact terms, and so on).

The correlation of measurement data in n series of measurement results makes the estimation of the statistical characteristics describing such a random sample difficult. The positive data correlation increases the variance of the arithmetic mean in relation to the series of uncorrelated results. If the normalized autocorrelation function of the observations is correlated positively and their variance is known then the impact of the correlation can be taken into consideration in the process of the estimation of the arithmetic mean variance in a computational manner. A significant hindrance to the assessment of the accuracy estimation process appears when the data variances and their autocorrelation function are unknown, since for the autocorrelation function with the relevant fragments of negative values the data correlation may decrease the variance of the arithmetic mean. The basic theoretical solutions of this problem are presented in literature, *e.g.* [1-2].

Due to the expanding range of the use of measuring data processing and the possibility of the use of new instruments and procedures, the problem of the assessment of the accuracy of

Article history: received on Feb. 1, 2012; received in revised form on Jul. 31, 2012; accepted on Sep. 20, 2012; available online on Dec. 15, 2012; DOI: 10.2478/v10178-012-0070-3.

the experimental determination of statistical characteristics (particularly the arithmetic mean) of correlated data was and still remains topical [3-16].

In this study the application of the conditional averaging of the correlated data with the normal distribution for the determination of the standard uncertainty of the arithmetic mean has been proposed. The results of theoretical research and the selected analysis experiments of physical signals have also been presented.

2. Assessment of the standard uncertainty of the arithmetic mean of correlated data

When compiling measurement data it is normally assumed that the observations taken with a sample interval Δt from the population (of a $x(t)$ signal) with the distribution $N(\mu_x,\sigma_x)$ create a n-element random time series (1) which meets the stationary and ergodicity conditions:

$$x_1(\Delta t), x_2(2\Delta t),..., x_i(i\Delta t),..., x_n(n\Delta t). \tag{1}$$

The commonly used estimator of the expected value μ_x determined on the basis of the data series (1) is the arithmetic mean calculated from the formula:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \ . \tag{2}$$

If the σ_x^2 observation variance is not known then it can be assessed with the unbiased variance estimate:

$$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \ . \tag{3}$$

When there is no data correlation then the standard uncertainty of the arithmetic mean can be calculated from the equation:

$$u_A(\bar{x}) = \frac{s_x}{\sqrt{n}} . \tag{4}$$

At the correlation of n data described with the normalized autocorrelation function $\rho_k = \rho(k\Delta t)$ the equivalents of the expressions (3) and (4) will have the form [3-4]:

$$s_{xk}^2 = \frac{n_e}{n_e - 1}\cdot\frac{n-1}{n}s_x^2, \tag{5}$$

and:

$$u_{Ak}(\bar{x}) = \frac{s_{xk}}{\sqrt{n_e}} = \frac{s_x}{\sqrt{n\dfrac{n_e - 1}{n-1}}}, \tag{6}$$

where:

$$n_e = \frac{n}{1 + 2\sum_{k=1}^{n-1}\left(1 - \dfrac{k}{n}\right)\rho_k} = \frac{n}{\lambda(n,\rho)} \tag{7}$$

is the effective (equivalent) number of the uncorrelated observations providing the same uncertainty as for n − correlated observations.

In the circumstances when the normalized autocorrelation function ρ_k of data is positive for any k values, then the coefficient $\lambda(n,\rho) > 1$, and this means that the positive data correlation decreases the accuracy of the estimation of the expected value.

For the n-element measurement data series described with the autocorrelation functions of exponential form according to the model:

$$R_x[k\Delta t] = \sigma_x^2 \rho_1^k, \quad k = 0,1,2,\ldots, \tag{8}$$

where: ρ_1 – correlation between the adjacent values of the data series, the coefficient $\lambda(n, \rho_1)$ can be presented by expression [1]:

$$\lambda(n, \rho_1) = \lambda = \frac{1+\rho_1}{1-\rho_1} - \frac{2\rho_1\left(1+\rho_1^n\right)}{n(1-\rho_1)^2} = \lambda_1 - \lambda_2 . \tag{9}$$

At the limit when $n \to \infty$:

$$\lambda(\infty, \rho_1) = \lambda_1 = \frac{1+\rho_1}{1-\rho_1}, \tag{10}$$

and the number of uncorrelated observations:

$$n_e = n\frac{1-\rho_1}{1+\rho_1}. \tag{11}$$

Substitution of the value λ with λ_1 in calculations means that the formula (7) is simplified to the form:

$$n_e = \frac{n}{1+2\sum_{k=1}^{n-1}\rho_k} = \frac{n}{\lambda_1}. \tag{12}$$

The error rate caused by this replacement is in percent:

$$\delta_\lambda = \frac{\lambda_1 - \lambda}{\lambda}100. \tag{13}$$

Figure 1 below shows the relationship between the error rate δ_λ and the function *n* for the value $\rho_1 = 0.2$ and $\rho_1 = 0.8$.

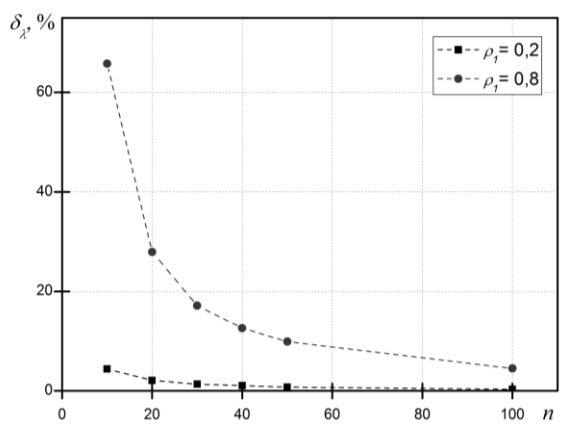


Fig. 1. Graph of the relationship $\delta_\lambda = f(n,\rho_1)$.

From Fig. 1 it can be seen that for high values of the correlation of adjacent elements of the data series the simplification of the relationship (9) to the form of (10) is burdened with a high error δ_λ for low values of *n*. For the condition when $\delta_\lambda < 10\%$ the number of results in the data series should not be less than 50.

3. Application of the conditional averaging for the assessment of standard uncertainty of the arithmetic mean

To calculate the uncertainty $u_{Ak}(\bar{x})$ with the use of conditional averaging of the random data time series with Gaussian distribution the following procedure can be applied:

1. Calculation of the estimate $\bar{x}$ (from the formula 2).
2. Calculation of the estimate s_x (from the formula 3).
3. Centering of the x_i data series and obtaining of the $\mathring{x}_i$ series as $\mathring{x}_i = x_i - \bar{x}_i$.
4. Calculation (on the basis of the $\mathring{x}_i$ series) of the estimate $\bar{x}_{w1}$ of the conditional values of the arithmetic mean:

$$\hat{E}\left(\mathring{x}_{i+1} \middle| \mathring{x}_i = x_p \right) = \bar{x}_{w1} = x_p \hat{\rho}_1 = \frac{1}{M} \sum_{m=1}^{M} \left(\mathring{x}_{i+1} \middle| \mathring{x}_i = x_p \right), \tag{14}$$

where: x_p – threshold value, M – number of averages.

The threshold x_p should be selected for the condition [17-18]:

$$x_p = \nu \cdot s_x, \tag{15}$$

where: ν – coefficient dependent on the assumed method of averaging.

5. After substitution of (14) and (15) to (11) and then to (5) and (6), new relationships for n_e^*, s_x^{2*} and u_A^* are obtained:

$$n_e^* = n \frac{x_p - \bar{x}_{w1}}{x_p + \bar{x}_{w1}}, \tag{16}$$

$$s_{xk}^{2*} = \frac{n-1}{n - \dfrac{x_p + \bar{x}_{w1}}{x_p - \bar{x}_{w1}}} s_x^2, \tag{17}$$

$$u_{Ak}^*(\bar{x}) = \frac{s_x}{\sqrt{n \dfrac{x_p(n-1) - \bar{x}_{w1}(n+1)}{(n-1)(x_p + \bar{x}_{w1})}}}. \tag{18}$$

For $n >> 1$ the last expression can be simplified and reduced to the form:

$$u_{Ak}^*(\bar{x}) \approx \frac{s_x}{\sqrt{n \dfrac{x_p - \bar{x}_{w1}}{x_p + \bar{x}_{w1}}}} = \frac{s_x}{\sqrt{n_e^*}}. \tag{19}$$

If $\bar{x} \neq 0$, then:

$$\bar{x}_{w1} = \bar{x} + \rho_1 (x_p - \bar{x}) \tag{20}$$

and the number of the uncorrelated observation will be:

$$n_e^* = \frac{x_p - \bar{x}_{w1}}{x_p + \bar{x}_{w1} - 2\bar{x}}. \tag{21}$$

In order to assess the degree of the correlation of the subsequently averaged signal fragments exceeding the set level x_p, the ratio $\tau_{km} / \bar{\tau}_{x_p}$ can be used (where τ_{km} is the maximum signal correlation interval and $\bar{\tau}_{x_p}$ is the average interval between the passes of x_p). It can be shown that for the model considered in this study of data correlation the subsequent passes of the level $x_p = \sqrt{2} \cdot \sigma_x$ are practically uncorrelated.

The relative uncertainty $\delta_{\bar{x}_{w1}}$ of estimate $\bar{x}_{w1}$ is determined by the ratio of standard uncertainty $\sigma_{\bar{x}_{w1}} = \sigma_x \sqrt{(1-\rho_1^2)/M}$ of conditional arithmetic mean M for the values of uncorrelated realizations to the conditional expected value $\bar{x}_{w1} = x_p \rho_1$ of the analyzed realizations.

The quality of estimate $\bar{x}_{w1}$, with specified length of the analyzed time series on the one hand requires a possibly large number M of conditionally averaged fragments of the time series in order to reduce the variance of the conditional arithmetic mean and to adequately decrease threshold x_p. On the other hand, the quality of averaging requires possibly large averaged values and an adequately high value of estimate $\bar{x}_{w1}$, which requires opposite action and increased threshold x_p. A compromise and optimum value of threshold x_p, which can be obtained from the minimum condition $\delta_{\bar{x}_{w1}}$, depends on the algorithm for determining $\bar{x}_{w1}$ and is found in the range $(1.4 \div 2.0)s_x$.

For $\bar{x} = 0$ and with lack of correlation of the signal realizations M subsequently averaged, the expression determining the relative uncertainty of the conditional assessment of the signal arithmetic mean $\bar{x}_{w1}$ can be depicted with the equation:

$$\delta_{\bar{x}_{w1}} = \frac{\sigma_{\bar{x}_{w1}}}{\bar{x}_{w1}} = \frac{\sigma_x \sqrt{\dfrac{1-\rho_1^2}{M}}}{x_p \cdot \rho_1} = \frac{1}{\nu \rho_1} \sqrt{\frac{1-\rho_1^2}{M}} \quad . \tag{22}$$

The graph of the relationship $\delta_{\bar{x}_{w1}} = f(\rho_1)$ for $\nu = \sqrt{2}$ and $M = 100$ is shown in Fig. 2. For $\rho_1 \geq 0{,}5$ the value $\delta_{\bar{x}_{w1}}$ does not exceed 2 percent.

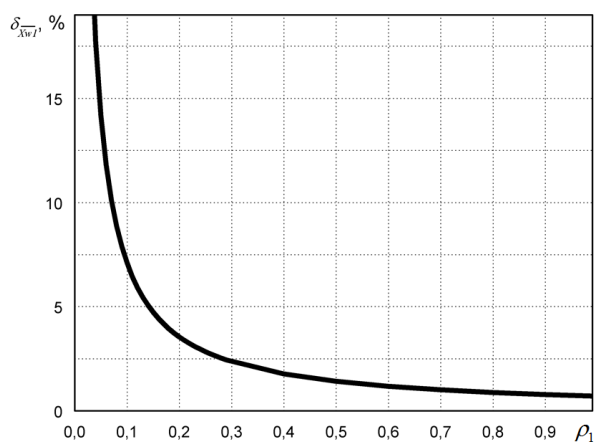
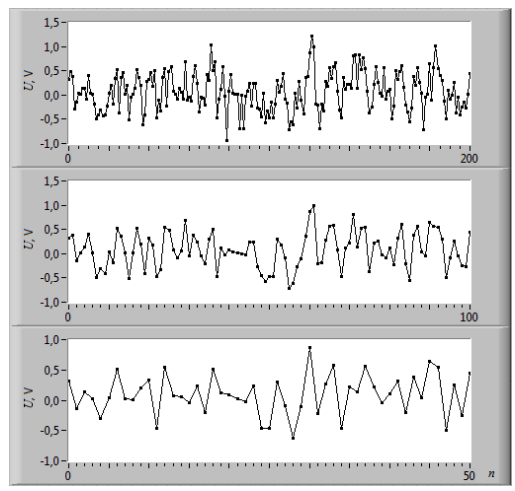


Fig. 2. The graph of the relationship $\delta_{\bar{x}_{w1}} = f(\rho_1)$ for $\nu = \sqrt{2}$ and $M = 100$.

4. Experimental studies

In the experiment the stationary random signal $x(t)$ of the $N(\mu_x, \sigma_x)$ distribution and exponential autocorrelation function was analyzed. The examined signal was obtained via processing of the output signal of a physical white noise generator within the band set for 0 – 25 kHz by the use of the inertial system of the first order with the time-constant $RC = 10^{-4}$ s. The first stage of the data analysis was the determination of the conditional estimate of the mean value $\bar{x}_{w1}$ with the use of the general relationship $\bar{x}_w(k\Delta t)$. This estimate was determined with the use of the RIGOL DS1062C digital oscilloscope for the threshold $x_p = 1.12\ \mathrm{V}\ (\nu = 2)$ and the number of averaging steps M = 256. The examples of records in the analyzed signal realizations $x(i\Delta t_p)$ for three sampling intervals: $\Delta t_{p1} = 50\,\mu$s, $\Delta t_{p2} = 100\,\mu$s and $\Delta t_{p3} = 200\,\mu$s, along with the graph of the conditional function of the mean value $\bar{x}_w = \mathrm{f}(k\Delta t_{p1})$ are shown in Fig. 3.

a)



b)

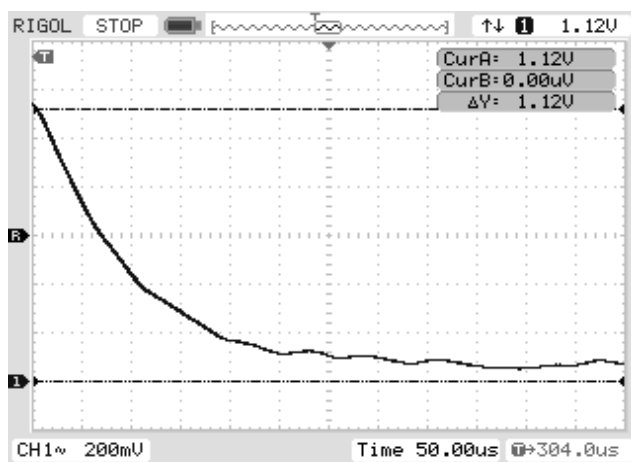


Fig. 3. Experimental characteristics: a) fragments of signal realizations $x(i\Delta t_p)$ from sampling intervals Δt_p equal subsequently to: Δt_{p1} = 50 μs, Δt_{p2} = 100 μs , and Δt_{p3} = 200 μs; b) function of the conditional mean value of the analyzed signal $\bar{x}_w = \mathrm{f}(k\Delta t_{p1})$.

In the second stage of the experiment carried out with a computer system, the mean value $\hat{\mu}_x = 0.073\,\mathrm{V}$ and the variance $\sigma_x^2 = 0.312\ \mathrm{V}^2$ of the signal $x(t)$ have been determined for long time observations ($T_0 = 100\,\mathrm{s}$). Then, 100 short realizations have been analyzed which included 200 signal samples taken with the sampling interval $\Delta t_{p1} = 50\mu s$. For each realization the arithmetic mean $\bar{x}_j$ and the variance $s_{x_j}^2$ were determined. Then the following experimental values have been determined: the arithmetic mean $\hat{\mu}_{x1}$ of the estimates $\bar{x}_j$, the signal variance s_{x1}^2, and the arithmetic mean variance $s_{\hat{\mu}_{x1}}^2$ (see the first column in Table 1). The experiment has been repeated 10 times (g = 10) for subsequent, different short 100-element realizations. On the basis of the statistics calculated in 10 repetitions (Table 1) the general arithmetic mean has been determined using the formula:

$$\hat{\mu}_x = \frac{\sum_{g=1}^{10} 100\hat{\mu}_{xg}}{10 \cdot 100} = 0.0729\ \mathrm{V} \tag{23}$$

and the general variance of the estimate of the arithmetic mean by using the formula:

$$s_{\hat{\mu}_x}^2 = \frac{\sum_{g=1}^{10} 100 s_{\hat{\mu}_{xg}}^2}{10 \cdot 100} + \frac{\sum_{g=1}^{10} 100\left(\hat{\mu}_{xg} - \hat{\mu}_x\right)^2}{10 \cdot 100} = 0.007 + 0.00006 = 0.0071\ \mathrm{V}^2 . \tag{24}$$

Table 1. The list of the calculation results.

g / Param.	1	2	3	4	5	6	7	8	9	10	Gen. param
$\hat{\mu}_{xg}$, V	0.0654	0.0854	0.0807	0.0761	0.0736	0.0787	0.0733	0.0719	0.0647	0.0595	$\hat{\mu}_x$ 0.0729
s_{xg}^2, V^2	0.3046	0.3069	0.3058	0,3103	0.3201	0.3074	0.3121	0.3109	0.3078	0.3061	$\hat{s}_x^2$ 0.3092
$s_{\hat{\mu}_{xg}}^2$, V^2	0.0056	0.0064	0.0084	0.0057	0.0077	0.0072	0.0060	0.0098	0.0066	0.0066	$s_{\hat{\mu}_x}^2$ 0.0070

The standard uncertainty is:

$$s_{\hat{\mu}_x} = \sqrt{s_{\hat{\mu}_x}^2} = \sqrt{0{,}0071} \approx 0.084\ \mathrm{V}. \tag{25}$$

With the use of the data conditional averaging, the effective number n_e^* of samples for $\Delta t_{p1} = 50\mu s$ has been calculated on the basis of data, according to Fig. 3b, by using the formula (21):

$$n_e^* = n\frac{x_p - \bar{x}_{w1}}{x_p + \bar{x}_{w1} - 2\bar{x}} = 200\frac{1.12 - 0.74}{1.12 + 0.74 - 2 \cdot 0.073} \approx 44$$

Next, the standard uncertainty of the arithmetic mean has been determined by the use of the formula:

$$u_{Ak}^*(\bar{x}) = \frac{S_x}{\sqrt{n_e^*}} = \frac{\sqrt{0.312}}{\sqrt{44}} \approx 0.084\,\mathrm{V}. \tag{26}$$

The signal analysis described herein has been made subsequently for longer sampling intervals $\Delta t_{p2} = 100\mu s$ and $\Delta t_{p3} = 200\mu s$, and the $s^2_{\hat{\mu}_x} = 0.0035\,\mathrm{V}^2$ as well as $s^2_{\hat{\mu}_x} = 0.0018\ \mathrm{V}^2$ values were obtained, respectively. The obtained experimental results $s^2_{\hat{\mu}_{xg}}$ and $s^2_{\hat{\mu}_x}$ of the assessment of the arithmetic mean variance are shown in Fig. 4. The graphs presented in Fig. 4 illustrate the decrease of the value of the arithmetic mean variance estimates at longer sampling intervals and lower correlation of the data correlated positively. The decrease of the variances of the variance estimates at the lowering of data correlation is also clear.

The standard uncertainty values determined on the basis of the comparative principles (*e.g.* the relationships (25) and (26)) for different sampling intervals and various data correlation levels are shown in Table 2.

On the basis of the presented result it can be stated that for relatively large data sets, for practical purposes there is sufficient conformity of the value of the standard uncertainty assessment with the use of the conditional mean and the classic method of the estimation of the standard deviation.

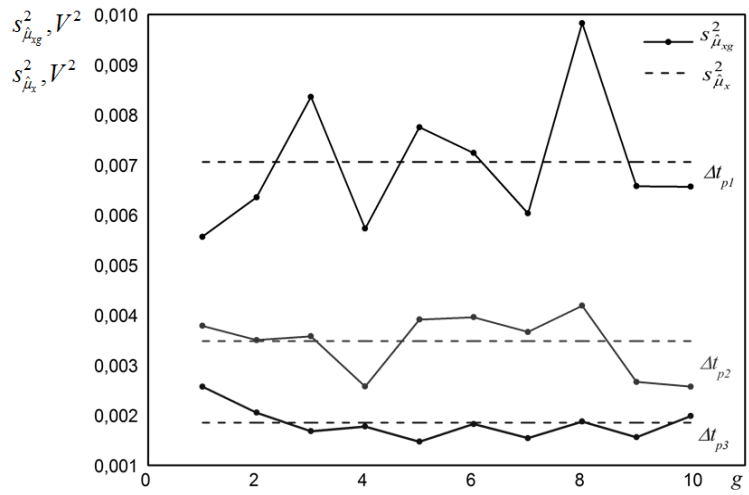


Fig. 4. Experimental values of the arithmetic mean variance.

Table 2. Experimental results of data analysis.

Assessment of the standard uncertainty	Standard uncertainty values for arithmetic mean [V]		
	Δt_{p1}	Δt_{p2}	Δt_{p3}
u^*_{Ak} [V] – determined from the formulas (21) and (19)	0.084 ($n^*_e = 44$)	0.057 ($n^*_e = 96$)	0.043 ($n^*_e = 169$)
$s_{\hat{\mu}_x}$ [V] – determined by experiment	0.084	0.059	0.043

5. Conclusion

For the assessment of the correlation of the measurement data series and the standard uncertainty of the correlated data arithmetic mean of the Gaussian distribution, conditional data averaging can be used. The method proposed herein can be particularly useful for high correlation values ($\rho_1 > 0.5$), and relatively high values of n ($n > 50$).

For the time series of data of limited length (*i.e.* low *n* value) the values $\bar{x}_{w1}$ of the estimate on the basis of formula (14) may not be stable in time (like the estimates of the autocorrelation function) because of the small number of occurring and conditionally averaged realizations.

The conditional averaging presented in this study for the basic model loses its advantages with regard to the assessment of correlation for insufficiently correlated data ($\rho_1 < 0.1$) due to the significant increase of the value of $\delta_{\bar{x}_{w1}}$. In such circumstances a beneficial solution may be given by the use of the modified algorithms of conditional averaging of measurement data [19].

Acknowledgements

This work was supported by the Ministry of Science and Higher Education, Poland (grant No. N N505 466038).

References

[1] Bartels J. (1935). Zur Morphologie geophysikalischer Zeitfunktionen. Sitz-Ber. *Preuß. Akad. Wiss*., 30, 502-522.

[2] Bayley G.V. & Hammersley G.M. (1946). The "effective" number of independent observations in an autocorrelated time-series. *J. Roy. Stat. Soc. Suppl.*, 8, 184-197.

[3] Bendat J.S., Piersol A.G. (2000). Random Data. Analysis and Measurement Procedures. *John Wiley & Sons*, New York.

[4] Box G.E.P., Jenkins G.M., Reinsel G.C. (1994). Time Series Analysis: Forecasting and Control. *Prentice Hall*, Englewood Cliffs.

[5] Dorozhovets M., Warsza Z. (2007). Evaluation of the uncertainty type A of autocorrelated measurement observations. *Measurement Automation and Monitoring*, 2, 20-24. (in Polish)

[6] Dorozhovets M. (2009). Influence of lack of a priori knowledge about autocorrelation functions of observations on estimation of their average value standard uncertainty. *Measurement Automation and Monitoring*, 55(12), 2-5. (in Polish)

[7] Freund R.J., Wilson W.J., (2006). Regression Analysis. Statistical Modeling of a Response Variable. *Elsevier*, Amsterdam.

[8] Kirkup L., B. Frenkel L. (2006). *An Introduction to the Uncertainty in Measurement.* Cambridge University Press, Cambridge.

[9] Leith C.E. (1973). The standard error of time-averaged estimates of climatic means. *J. Appl. Meteorol*., 12, 1066-1069.

[10] Şen Z. (1998). Small sample estimation of the variance of time-averages in climatic time series. *Int. J. Climatol*., 18, 1725-1732.

[11] Witt T.J. (2007). Using the autocorrelation function to characterize time series of voltage measurements. *Metrologia*, 44, 201-209.

[12] Zhang N.F. (2008). Allan variance of time series models for measurement data. *Metrologia*, 45, 549-561.

[13] Zhang N.F. (2006). Calculation of the uncertainty of the mean of autocorrelated measurements. *Metrologia*, 43, 276-281.

[14] Zięba A. (2010). Effective number of observations and unbiased estimators of variance for autocorrelated data – an overview. *Metrol. Meas. Syst.*, 17(1), 3-16.

[15] Zięba A., Ramza P. (2011). Standard deviation of the mean of autocorrelated observations estimated with the use of the autocorrelation. *Metrol. Meas. Syst.*, 18(4), 529-542.

[16] Kowalczyk A., Szlachta A., Hanus R. (2011). Application of correlation interval to determination of standard uncertainty of arithmetic mean for correlated data. *Measurement Automation and Monitoring*, 57(12), 1549-1551. (in Polish)

[17] Mirskii, G.J. (1971). Instrumental determination of the characteristics of random processes. *Energiya*, Moscow. (in Russian)

[18] Kowalczyk A., Szlachta A. (2010). The application of conditional averaging of signals to obtain the transportation delay. *Electrical Review*, 86(1), 225-228. (in Polish)

[19] Hanus R., Szlachta A., Kowalczyk A., Petryka L., Zych M. (25-28 March, 2012). Radioisotope Measurement of Two-Phase Flow in Pipeline Using Conditional Averaging of Signal. In *Proc. IEEE Mediterranean Electrotechnical Conference MELECON 2012*. IEEE Press, 144-147.