UDK 519.222

Warsza Z. L., PhD.

EVALUATION OF THE TYPE A UNCERTAINTY IN MEASUREMENTS WITH AUTOCORRELATED OBSERVATIONS

Expanding of the application range of the GUM uncertainty type A evaluation method to the case of regularly sampled autocorrelated observations is described. As the first step is the previous identification and cleaning of the raw sample data from the regularly variable components. Then formulas for standard deviation of the mean value are expressed with the use of the so-called "effective number" of observations. This quantity depends on number of observations and on the sample autocorrelation function and allows calculating the expanded uncertainty due GUM recommendations. Given is method of estimation of autocorrelation function for the sample data. Considerations are illustrated by examples.

Keywords: *measurement, measurement uncertainty, autocorrelation function, random distribution.*

Варша З.Л., канд. техн. наук

ОЦЕНКА НЕОПРЕДЕЛЕННОСТИ ИЗМЕРЕНИЙ ТИПА А С АВТОКОРРЕЛЯЦИЕЙ НАБЛЮДЕНИЙ

Представлено расширение области ГУМ неопределенности типа А. Описан метод оценки для случая регулярных автокорреляционных наблюдений. В качестве первого шага используется первичная идентификация и очистка первичных данных выборки из различных переменных компонентов. Тогда формулы для стандартного отклонения от среднего значения выражаются с помощью так называемого "эффективного числа" наблюдений. Эта величина зависит от числа наблюдений, и автокорреляционная функция образца позволяет рассчитать расширенную неопределенность благодаря рекомендациям ГУМ. Приводится метод оценки автокорреляционной функции для выборки данных. Соображения иллюстрируются примерами.

Ключевые слова*: измерение, погрешности измерений, автокорреляционная функция, случайное распределение.*

Варша З.Л., канд. техн. наук

ОЦІНКА НЕВИЗНАЧЕНОСТІ ВИМІРЮВАНЬ ТИПУ А З АВТОКОРЕЛЯЦІЄЮ СПОСТЕРЕЖЕНЬ

Наведено розширення області ГУМ невизначеності типу А. Описано метод оцінки на випадок регулярних автокореляційних спостережень. В якості першого кроку використовується попередня ідентифікація та очищення первинних даних виборки з різних змінних компонентів. Тоді формули для стандартного відхилення від середнього значення виражаються за допомогою так званого "ефективного числа" спостережень. Ця величина залежить від числа спостережень і автокореляційна функція зразка дозволяє розрахувати розширену невизначеність завдяки рекомендаціям ГУМ. Наводиться метод оцінки автокореляційної функції для вибірки даних. Міркування ілюструються прикладами.

Ключові слова: *вимірювання, похибки вимірювань, автокореляційна функція, випадковий розподіл.*

Introduction. The international guide about evaluation of uncertainty in measurement known under acronym GUM covers only measurement of the variable with randomly distributed but of no related statistically observations (i.e. without autocorrelation of their data). Then use of GUM is limited in many types of measurements. In particular, it was not established how to estimate uncertainty in measurement of the processes variable in time or space. This paper will synthetically discuss the results of some Polish work on the determination of measurement uncertainty of regularly sampled measured, covered in detail in publications [1–7] and their bibliographies.

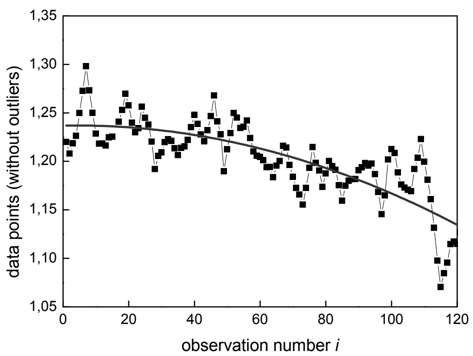
This includes preparing of the raw data sample to farer calculations by removal from it a priori unknown regularly variable components (i.e. "cleaning of the raw sample"),the estimation of sample proper uncertainty type A for know autocorrelation function ρ and is given how to find from the sample data the estimator r of the autocorrelation function.

"Cleaning" The Raw Data. The process of collecting measurement observations is now usually automated. The values of the signal or the output readings are discrete as a result of sampling the input analog signal with properly chosen frequency and A/D conversion. The dispersion of the "raw" values of measurement observations is caused by reason of both random and determined type, as a result of changes of measurand itself, changes of the internal parameters of the measuring circuit and environmental conditions. It is different for each of the samples taken at different times, measured by different instruments, and even by the same device over its lifetime. Random changes of the observations can be stationary and no stationary. There are short-term noise (called outliers), which prior to the assignment of values and the uncertainty of the measurements should be identified and eliminated from the raw data. Changes in the form of a regular no periodic component, i.e. the trend and periodic components affect the shape of the histogram of the sample, the mean value and uncertainty of type A. So there are also undesirable.

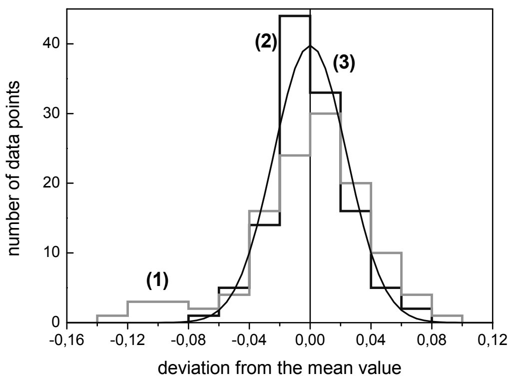
Example 1. Fig. 1, a shows a series of successive $n = 121$ "raw" results of measurement observations (black squares), obtained by regular sampling of the tested process [3]. You can see that the sample contains non-linear decreasing trend, there are not noticeable periodic components, and rather a small laceration of the data graph indicates that this data can be auto correlated. After removing one outlier, using LSM (least squares) was determined the polynomial

$$y(i) = (-7{,}65 \cdot 10^{-7})\, i^2 + (6{,}6 \cdot 10^{-5}) i + 0{,}034,$$

modeling the trend and was removed from the sample data, assuming that it passes through the average value. In Fig. 1, b a histogram of the raw sample is given – (1) and histogram after removal of the trend – (2) and a matching criterion $\chi 2$ its Gaussian *pdf* function – (3).



a



b

Fig. 1. An example of the measurement data sample collected sequentially by regular sampling: deviations from the mean value of the raw data and their systematic regularly variable component – drift identified by the least squares method (a);
distributions: of the raw data (1), after subtraction of the drift (2) and Gaussian pdf curve fitted to that data (3) (b)

Trend causing asymmetry and the existence of a large "tail" on the left of histogram (1), which does not meet the criterion of $\chi 2$, and with trend uncertainty u_A is about 64% higher than u_A=0,00219 without it. Purification of the raw data significantly reduced u_A. For n=120 standard deviation $s(u_A)/s \approx 8{,}5$ %.

The Uncertainty of Mean Value for Correlated Observations. Presented is a brief description of the problem. 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

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

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

The relationship between standard deviation $\sigma(\bar{x})$ of the mean value and σ of the individual correlated observation results from the variance of the sum of random variables [1, 2, 4]

$$\frac{\sigma(\bar{x})}{\sigma} = \frac{1}{\sqrt{n_{eff}}}. \qquad (2)$$

Where

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

For the statistically independent observations $\rho_k \to 0$ (for $k \geq 0$), consequently $D_\rho = 0$ and formula (2) passes to the known relation

$$\sigma(\bar{x})\big/\sigma = 1\big/\sqrt{n}\ .$$

In opposite, when the observations are fully correlated (closely linked), i.e. $\rho_k \to 1$, with (3) results

$$D_\rho \to \frac{2}{n}\sum_{k=1}^{n-1}\left(n-k\right)\cdot 1 = n-1. \qquad (4)$$

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

The value of n_{eff} is needed to properly estimate the standard deviation of samples of correlated observations [4–6].

$$s_a(\bar{x}) = \frac{s_a(x_i)}{n_{eff}} = \sqrt{\frac{1}{n\ (n_{eff}-1)}\sum_{i=1}^{n}(x_i-\bar{x})^2}, \qquad (5)$$

or

$$s_a\left(\bar{x}\right) = k_b s\left(\bar{x}_i\right),$$

where

$$s^2\left(x_i\right) = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2,$$

$$s\left(\bar{x}_i\right) = \frac{s\left(x_i\right)}{\sqrt{n}},$$

$$s_a(x_i) = k_a\ s(x_i),$$

$$k_a = \sqrt{\frac{n_{eff}(n-1)}{n(n_{eff}-1)}} \approx 1;\ k_b = \sqrt{\frac{n-1}{n_{eff}-1}}.$$

For the autocorrelation data can be also used the effective number degrees of freedom ν_{eff} defined approximately [5–7] as

$$\boxed{\nu_{eff} \cong \frac{n}{1+2\sum_{k=1}^{n-1}\rho_k^2} - 1.} \qquad (6)$$

But $\nu_{eff} \neq n_{eff}-1$ [5–7]. The relative dispersion of the standard deviation is

$$\frac{u(s_a)}{s_a} = \frac{u(s_a(\bar{x}))}{s_a(\bar{x})} \cong \frac{1}{\sqrt{2\nu_{eff}}}. \qquad (7)$$

Estimator of Autocorrelation Function of Measurement Data Sample. Autocorrelation function is usually not known and needs to estimate from the measurement data. The most commonly used and implemented in the programs is its form

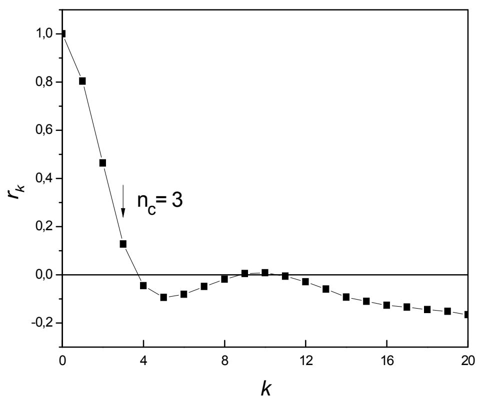
$$r_k = \frac{\sum_{i=1}^{n-k}\left(x_i-\bar{x}\right)\left(x_{i+k}-\bar{x}\right)}{s^2(q_i)}. \qquad (8)$$

Estimate r_k (Fig. 2,a) has two qualitatively different parts. For small distances k is the falling edge, which is contained real information about the autocorrelation function. The remainder tail is the image of a rather large fluctuations of the correlated noise.

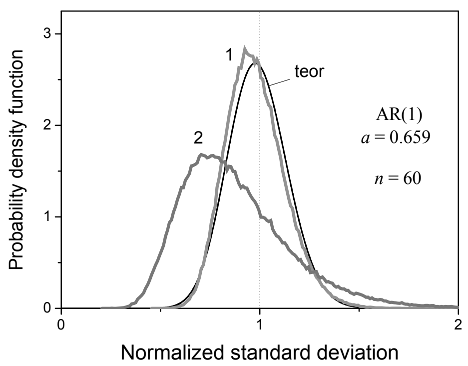
Replace the function ρ_k by its estimate r_k in formula (3) gives of estimate of the effective number of observations n_{eff} not of the satisfied properties. The reason is the influence of autocorrelation function of the tail. According Zięba [5–7] summation in (8) can be reduced to only a few initial estimate of r_k elements, i.e.

$$\hat{n}_{eff} = \frac{n}{1+2\sum_{k=1}^{n_c}\left(1-\frac{k}{n}\right)r_k}. \qquad (9)$$

The border value n_c is determined by the last non-zero element of r_k estimate before its first passage through zero (FTZ method – called from the first transit through zero).



a



b

Fig. 2. The initial part of the estimate $\{r_k\}$ of autocorrelation function ρ_k computed from data of Fig. 1, a (a).
Probability density functions 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)

For example, this value for the curve on Fig. 2, a is n_c=3. FTZ method is valid only for positive correlations. Fig. 2, b [6], [4] shows two examples of probability distributions of the estimators obtained by Monte Carlo method.

Curves 1 and 2 are derived from MC simulations and relate, respectively, s_a/σ, and $s_a(\bar{x})/\sigma(\bar{x})$. Theoretical curve teor has been calculated theoretically using the model (9) for $\nu_{\text{eff}} = 22{,}7$.

The distribution marked as *theor* on fig. 2, b 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 ν_{eff}. Confirmed that the sum of squared differences (5b) is susceptible to distribution of $\chi 2$ with $\nu_{eff} \equiv \nu$.The simulation studies using MC [7] show that in this cause estimator $\hat{n}_{eff}$ reduce the negative bias of the mean value $\bar{x}$. Obtained value of $\hat{n}_{eff}$ is used for calculations of $s_a(x_i)$ and $s_a(\bar{x})$ by (5d) and (5).

Example 2

Let us calculate 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 from Example 1 without trend and for the estimator r_k of their autocorrelation function for n_c=3 according to Fig. 2,a. A large value of r_k=0.81 confirms autocorrelation of these data. Without considering it the standard deviation $s(q_i)$=0,0241. At this value of r_k (9) implies the estimate n_{eff}=32,1, and from (5) the coefficient k_a=1.012 and k_b=1,96. Correlation 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 (6) and (7): $u(s_a)/s \approx 11$ %.

Summary and Conclusions. Discussed issues arose from the purpose of determining the measurement uncertainty for the measurement data obtained at a sampling. This paper summarizes the results of research about cleaning such raw data from their regular components [3] and the extension of calculating the uncertainty by the method type A of GUM to the auto correlated data [1,2] – [4–7]. The developed method is easy to use in the practice of measurement and can provide a basis to adopting it into the GUM upgrading. Conclusions:

• For a limited time to collect measurement observations, a reduction of measurement uncertainty by increasing the sample size by increasing the sampling rate is unreliable, because it leads to the need to reflect the impact of autocorrelation function of observation.

• Before calculating the uncertainty $u_A(x)$ must first be used appropriate computational methods to identify and remove from the raw results the systematic no periodic and periodic components.

• For such a cleaned values of the observations one needs to know or estimate the autocorrelation function. This function causes a significant increase in uncertainty u_A 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.

• 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.

• Evaluation of the uncertainty u_A for auto correlated data discussed here, concerns as in GUM, for the model of normal distribution. For other distributions, this method requires further investigation.

References

1. Dorozhovetz M. Proposals for the extension methods of determining the uncertainty of measurements by GUM guide / M. Dorozhovetz, Z. L. Warsza. *PAR* No. 1. – 2007. – P. 16–25 [In Polish].

2. Warsza Z. L. Uncertainty type a evaluation of autocorrelated measurement observations. Biuletyn WAT (Military Technical Academy) / Z. L. Warsza, M. Dorozhovets. – Warszawa : – Vol. LVII. – No 2. – 2008. – P. 143–152 [In Polish].

3. Warsza Z. L. Eliminacja wpływu nieznanych a priori składowych systematycznych na niepewność typu A pomiarów o równomiernym próbkowaniu. (Elimination of influence of unknown a priori systematic components on the uncertainty type A in measurement of regular sampling). PAR Pomiary Automatyka Robotyka (Measurements Automation Robotics) / Z. L. Warsza, J. Korczynski. – 2008. – No 2. – P. 5–13 [in Polish].

4. Warsza Z. L. Niepewność typu a pomiaru o obserwacjach samoskorelowanych (Measurement uncertainty type A of autocorrel-ated observations) / Z. L. Warsza, A. Zieba. PAK Pomiary Automatyka Kontrola. – No. 2. – 2012. –P. 157–162 [in Polish].

5. Zięba A. Effective number of observations and unbiased estimators of variance for auto correlated data – an overview / A. Zieba. Metrology& Measurement Systems, 17. – 2010. P. 3–16 [in Polish].

6. Zieba A. Niepewność pomiaru dla ciągu obserwacji samo skorelowanych (Uncertainty of measurement of the auto correlated observation) / A. Zieba. Monograph: Niepewność pomiarów w teorii i praktyce. Pracazbiorowa. (The uncertainty of measurement in theory and practice.Collective work). GUM. – Warsaw: – 2011. – P. 109–118 [in Polish].

7. Zięba A. Standard deviation of the mean of auto correlated observations estimated with the use of the autocorrelation function estimated from the data / A. Zięba, P. Ramza. Metrology & Measurement Systems, 18. – 2011. P. 529–534 [in Polish].

Received 15.05.2012

Warsza Zygmund Lech,
asoc. prof., PhD. El. Eng.
Former President of Polish Metrological Society